

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Takanori OKURA  
Kakuji TORIGOE  
Masahi KURIMOTO

(ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF INDUCING THE PRODUCTION OF INTERFERON- $\gamma$

(iii) NUMBER OF SEQUENCES: 35

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: BROWDY AND NEIMARK  
(B) STREET: 419 Seventh Street, N.W., Suite 300  
(C) CITY: Washington  
(D) STATE: D.C.  
(E) COUNTRY: USA  
(F) ZIP: 20004

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: Patent In Release #1.0, Version #1.30

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: JP 185,305/96  
(B) FILING DATE: 27-JUN-1996

(vii) ATTORNEY/AGENT INFORMATION:

(A) NAME: BROWDY, Roger L.  
(B) REGISTRATION NUMBER: 25,618  
(C) REFERENCE/DOCKET NUMBER: OKURA=1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 202-628-5197  
(B) TELEFAX: 202-737-3528

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 157 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn  
1 5 10 15  
Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp  
20 25 30  
Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile  
35 40 45  
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile  
50 55 60  
Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile  
65 70 75 80  
Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys  
85 90 95  
Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys  
100 105 110  
Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu  
115 120 125

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu  
 130 135 140  
 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp  
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1120 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: liver
- (ix) FEATURE:
  - (A) NAME/KEY: 5'UTR
  - (B) LOCATION: 1..177
  - (C) IDENTIFICATION METHODS: E
  - (A) NAME/KEY: leader peptide
  - (B) LOCATION: 178..285
  - (C) IDENTIFICATION METHODS: S
  - (A) NAME/KEY: mat peptide
  - (B) LOCATION: 286..756
  - (C) IDENTIFICATION METHODS: S
  - (A) NAME/KEY: 3'UTR
  - (B) LOCATION: 757..1120
  - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCCTGGACAG	TCAGCAAGGA	ATTGTCTCCC	AGTGCATTT	GCCCTCTGG	CTGCCAACTC	60
TGGCTGCTAA	AGCGGCTGCC	ACCTGCTGCA	GTCTACACAG	CTTCGGGAAG	AGGAAAGGAA	120
CCTCAGACCT	TCCAGATCGC	TTCCTCTCGC	AACAAACTAT	TTGTCGCAGG	AATAAG	177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG						225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met						
-35	-30	-25				
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC						273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn						
-20	-15	-10	-5			
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA						321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile						
1	5	10				
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT						369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro						
15	20	25				
CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG						417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg						
30	35	40				
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG						465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met						
45	50	55	60			
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT						513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys						
65	70	75				
GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC						561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile						
80	85	90				
AAG GAT ACA AAA AGT GAC ATC ATA TTC CAG AGA AGT GTC CCA GGA						609

Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly  
95 100 105

CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT	657
His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe	
110 115 120	
CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA	705
Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys	
125 130 135 140	
GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA	753
Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu	
145 150 155	
GAC TAGCTATTAA AATTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA	806
Asp	
GCCCTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAGGTCA GGTGTTCAAG ACCAGCCTGA	866
CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG	926
CATGCCCTCA ATCCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG	986
GTAGAGGTTG TGGTGAGCCG AGATTGCACC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA	1046
AACCTCCATCT CAAAAAATAA AATAAAATAA TAAACAAATA AAAAATTCTAT AATGTGAAAA	1106
AAAAAAAAAA AAAAA	1120

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 135 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
  - (A) NAME/KEY: exon
  - (B) LOCATION: 1..135
  - (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA	47
Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser	
-5 1 5 10	
GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT	95
Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn	
15 20 25	
CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G	135
Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp	
30 35 40	

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 134 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta

- (ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..134  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC	47
Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser	
40 45 50 55	
CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT	95
Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile	
60 65 70	
TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG	134
Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys	
80 85	

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 87 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (F) TISSUE TYPE: placenta
- (ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..87  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAATAAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG	50
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val	
-35 -30 -25	
GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G	87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 12 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:  
 (A) ORGANISM: human  
 (F) TISSUE TYPE: placenta
- (ix) FEATURE:  
 (A) NAME/KEY: exon  
 (B) LOCATION: 1..87  
 (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CT GAA GAT GAT G  
Ala Glu Asp Asp Glu  
-10

12

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2167 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: human  
(F) TISSUE TYPE: placenta

(ix) FEATURE:  
(A) NAME/KEY: exon + 3'UTR  
(B) LOCATION: 1..2167  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85 90 95 100	
TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA	96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
105 110 115	
TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
120 125 130	
CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT	192
Leu Phe Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
135 140 145	
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTCA TGCCGGCGC	246
Ile Met Phe Thr Val Gln Asn Glu Asp	
150 155	
AGTGGCTCAC GCCTGTAATC CCAGCCCTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG	306
TCAGGTGTC AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
AAAATTAGCT GAGTGTAGTG ACCCATGCC CTCATCCCAG CTACTCAAGA GGCTGAGGCA	426
GGAGAACATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTGAG CCGAGATTGC ACCATTGCGC	486
TCTAGCCTGG GCAAAACAG CAAAACTCCA TCTCAAAAAA TAAATAAAAT AAATAACAA	546
ATAAAAAAATT CATAATGTGA ACTGTCGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAAGT CTTGAAAAAC ATCATTAAAGA	666
ATGAAATGAA CTTTCACAAA AGCAACAAA CAGACTTTCC CTTATTTAAG TGAATAAAAT	726
AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTGA AAACATTCTA CATTGTTAAT	786
TGGCATATTA ATTATACTTA ATATAATTAT TTTAAATCT TTTGGTTAT TAGTCCTAAT	846
GACAAAAGAT ATTGATATTT GAACTTCTA ATTTTAAGA ATATCGTAA ACCATCAATA	906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TTCCCTAAA GTCACTATAT	966
TTTTAAATT CAGTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
CAACTGACAT TCATCCTAGG AATGCAAAGA TGGTTAATA TCCTAAAATC AATTAACATA	1086
ACATACTATA TTAATAAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAT	1146
TCCTGGTTA CACTATCATC TCAATAGATG CAGAAAAGC ATTTGACAAA ATCCAATTCA	1206
TAATAAAAAT TCTCAAACCTT GAAAGAGAAC ATCATAAAAGG CATCTATGAA AACCTACAG	1266
CTAATATCAT ACTAACGAT GAAAAACTGA ATTATTTAC CCTAAGATCA AGAATAATGC	1326
AAGCATGTCA GCTCTTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
CCATACAATA AATAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTT ATTGAGAC	1446
AACATGGTC TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT	1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAATA CATTGAAGGC TGGGCTCAGT	1566
GGAGATGGCA TGTACCTTTC GTCCCAGCTA CTTGGGAGGC TGAGGTAGGA GGATCACTG	1626
AGGTGAGGAG TTTGAGGCTA TAGTGAATG TGATCTTGC TGTGAATAGC CACTGCACTC	1686
GAGCCTAGGC AACAAAGTGA GACCCCGTCT CCAAAAAAAA AAATGGTATA TTGGTATTTC	1746
TGTATATGAA CAATGAATGA TCTGAAAACA AGAAAATTCC ATTACACGATG GTATTAACAA	1806
AATAAAAATAC AAATAAAATT AGCAAAATAA TTATAAAACT TGTACATCGA AAATTCAAA	1866

GCACCTCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA 1926  
 TTGGAAAATT CATTCAATAT TGTTAACATA ACAATTGTCC CCAAATTGAT GCATGCATTTC 1986  
 AATTAGTCT TCATCAAAAT TCCAGCAGGG TTTTGCGAGA AATTGACAAG CTGTACCCAA 2046  
 AATGTATATG GAAATGAAAA GACCCAGAAG AGCAAATAAT TTTTAAAAAA CAAAGTTGGA 2106  
 AAACCTTTAC TTCCTTAATT TAAAACCTAC TATAAACCTA AAGTTATCAA GACCATTAG 2166  
 T 2167

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1334 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION: 1..1334
  - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATTTTTT	TAATTCGAA	ACATAGAAAT	GACTAGCTAC	TTCTTCCCCT	TCTGTTTAC	60
TGCTTACATT	GTTCCGTGCT	AGTCCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTCACCTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTT	TAGTTGGGGT	AAAAAATTGG	180
ATACAATAAG	ACATTGCTAG	GGGTCTATGCC	TCTCTGAGCC	TGCCTTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTAAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTTACAGCTG	AAAATGCTGA	TAGTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TACTTGGGAG	GCTCAAGCAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
TGTGATCGTA	CCTGTGAATA	GCCACTGAC	TCCAGCCTGG	GTGATATACA	GACCTTGCT	480
CTAAAAATTAA	AAAAAAAAAA	AAAAAAACCC	TTAGGAAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCCAAA	ACATGAATT	CAAATATCAA	AGTTAGGCTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTTAA	AATACTGAAT	ACTTACCTTA	ACATATATT	AAATATTTT	ATTTAGCATT	660
TAAAAGTAA	AAACAACTTT	TTAGAATTCA	TATCTTTAA	ATACTAAAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACTGTGG	GGTTGTTTGT	TTGTTTGAGA	TGCAGTTCA	780
CTCTGTCAACC	CAGGCTGAAG	TGCAGTGCAG	TGCAGTGGTG	TGATCTCGGC	TCACTACAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	GCCTCAGTCT	CCCGAGTAGG	TGGGATATACA	900
GGCATGCACC	ACTTACACCC	GGCTAATT	TGTATTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTGGCCAG	GCTGGTCTCA	AAACCCCTAAC	CTCAAGTGAT	CTGCCTGCCT	CAGCCTCCCCA	1020
AAACAAACAA	CAACCCCCACA	GTTTAAATATG	TGTTACAAACA	CACATGCTGC	AACTTTTATG	1080
AGTATTTAA	TGATATAGAT	TATAAAAGGT	TGTTTTAAAC	TTTTAAATGC	TGGGATTACA	1140
GGCATGAGCC	ACTGTGCCAG	GCCTGAACGTG	TGTTTTAA	AATGTCTGAC	CAGCTGTACA	1200
TAGTCTCCTG	CAGACTGGCC	AAGTCTCAA	GTGGGAACAG	GTGTATTAAG	GAECTATCCTT	1260
TGGTTAAATT	TCCGCAAATG	TTCCTGTGCA	AGAATTCTTC	TAACTAGAGT	TCTCATTAT	1320
TATATTTATT	TCAG					1334

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4773 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
- (ix) FEATURE:

(A) NAME/KEY: intron  
 (B) LOCATION: 1..4773  
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTAAGACTGA	GCCTTACTTT	GTTTCAATC	ATGTTAATAT	AATCAATATA	ATTAGAAATA	60
TAACATTATT	TCTAATGTTA	ATATAAGTAA	TGTAATTAGA	AAACTCAAAT	ATCCTCAGAC	120
CAACCTTTG	TCTAGAACAG	AAATAACAAG	AAGCAGAGAA	CCATTAAGT	GAATACTTAC	180
AAAAAATTAT	CAAACCTTT	ACCTATTGTG	ATAATGATGG	TTTTCTGAG	CCTGTCACAG	240
GGGAAGAGGA	GATACAACAC	TTGTTTATG	ACCTGCATCT	CCTGAACAAT	CAGTCTTAT	300
ACAAATAATA	ATGAGAATA	CATATGTGAG	TTATACATT	AAGAATAACA	TGTGACTTTC	360
CAGAATGAGT	TCTGCTATGA	AGAATGAAGC	TAATTATCCT	TCTATATTTC	TACACCTTG	420
AAAATTATGA	TAATATTTA	ATCCCTAGTT	GGTTTGTG	TGATCCTTAG	CCTAAGTCTT	480
AGACACAAGC	TTCAGCTTC	AGTTGATGTA	TGTTATTTT	AATGTTAATC	TAATTGAATA	540
AAAGTTATGA	GATCAGCTGT	AAAAGTAATG	CTATAATTAT	CTTCAAGCCA	GGTATAAAGT	600
ATTTCTGGCC	TCTACTTTT	CTCTTATT	CTCCATTATT	ATTCTCTATT	ATTTTCTCT	660
ATTCCTCCA	TTATTGTTAG	ATAAACACAA	ATTAACTATA	GCTACAGACT	GAGCCAGTAA	720
GAGTAGCCAG	GGATGCTTAC	AAATTGGCAA	TGCTTCAGAG	GAGAATTCCA	TGTCAATGAA	780
ACTCTTTG	AGTGGAGATT	TGCCAATAAA	TATCCGCTT	CATGCCACC	CAGTCCCCAC	840
TGAAAGACAG	TTAGGATATG	ACCTTAGTGA	AGGTACCAAG	GGGCAACTTG	GTAGGGAGAA	900
AAAAGCCACT	CTAAAATATA	ATCCAAGTAA	GAACAGTGCA	TATGCAACAG	ATACAGCCC	960
CAGACAAATC	CCTCAGCTAT	CTCCCTCCAA	CCAGAGTGCC	ACCCCTTCAG	GTGACAATT	1020
GGAGTCCCCA	TTCTAGACCT	GACAGGCAGC	TTAGTTATCA	AAATAGCATA	AGAGGCCTGG	1080
GATGGAAGGG	TAGGGTGGAA	AGGGTTAACG	ATGCTGTTAC	TGAACAAACAT	AATTAGAAGG	1140
GAAGGAGATG	GCCAAAGCTCA	AGCTATGTGG	GATAGAGGAA	AACTCAGCTG	CAGAGGAGA	1200
TTCAGAAACT	GGGATAAGTC	CGAACCTACA	GGTGGATTCT	TGTTGAGGG	GACTGGTGAA	1260
AATGTTAAGA	AGATGGAAT	AATGCTTGGC	ACTTAGTAGG	AACTGGGCAA	ATCCATATT	1320
GGGGGAGCCT	GAAGTTTATT	CAATTGAT	GGCCCTTTA	AATAAAAAGA	ATGTGGCTGG	1380
CGCTGGTGGC	TCACACCTGT	AATCCCAGCA	CTTGGGAGG	CCGAGGGGGG	CGGATCACCT	1440
GAAGTCAGGA	GTTCAGAACCC	AGCCTGACCA	ACATGGAGAA	ACCCCATCTC	TAECTAAAAT	1500
ACAAAATTAG	CTGGGCGTGG	TGGCATATGC	CTGTAATCCC	AGCTACTCGG	GAGGCTGAGG	1560
CAGGAGAACATC	TTTGAAACCC	GGGAGGCAGA	GGTTGCGATG	AGCCTAGATC	GTGCCATTGC	1620
ACTCCAGCCT	GGCAGAACAAAG	AGCAAAACTC	GGTCTCAAAA	AAAAAAAAAAA	AAAAGTGAAA	1680
TTAACCAAAG	GCATTAGCTT	AATAATTAA	TACTGTTTT	AAGTAGGGCG	GGGGGTGGCT	1740
GGAAGAGATC	TGTGTAATG	AGGGAAATCTG	ACATTTAACG	TTCATCAGCA	TCATAGCAA	1800
TCTGCTTCTG	GAAGGAACCT	AATAAATTAT	AGTTGGAGGG	GGGGAGAGAG	TGAGGGGTGG	1860
ACTAGGACCA	TTTTAGGCC	TTGTCCTTAA	TCCCTTTCC	TGCCACTAAT	AAGGATCTT	1920
GCAGTGGTTA	AAAAAGTGGC	CTAGGTTCTA	GATAATAAGA	TACAACAGGC	CAGGCACAGT	1980
GGCTCATGCC	TATAATCCC	GCACTTGGG	AGGGCAAGGC	GAGTGTCTCA	CTTGAGATCA	2040
GGAGTTCAAG	ACCAGCCTGG	CCAGCATGGC	GATACTCTGT	CTCTACTAAA	AAAATACAA	2100
AAATTAGCCA	GGCATGGTGG	CATGCACCTG	TAATCCCAGC	TACTCGTGAG	CCTGAGGCAG	2160
AAGAATCGCT	TGAAACCAGG	AGGTGTAGGC	TGCACTGAGC	TGAGATCGCA	CCACTGCAC	2220
CCAGCCCTGG	CGACAGAAATG	AGACTTGTG	TCAAAAAAAAG	AAAAGATAAC	AACAGGCTAC	2280
CCTTATGTGC	TCACCTTCA	CTGTTGATTA	CTAGCTATAA	AGTCTCTATAA	AGTTCTTGG	2340
TCAAGAACCT	TGACAACACT	AAGAGGGATT	TGCTTTGAGA	GGTTACTGTC	AGAGTCTGTT	2400
TCATATATAT	ACATATACAT	GTATATATGT	ATCTATATCC	AGGCTTGGCC	AGGGTTCCCT	2460
CAGACTTCC	AGTGCACTTG	GGAGATGTTA	GGTCATATC	AACTTTCCCT	GGATTCAAGAT	2520
TCAACCCCTT	CTGATGTAAA	AAAAAAAAG	AAAAAGAAAG	AAATCCCTT	CCCCTTGGAG	2580
CACTCAAGTT	TCACCCAGGTG	GGGCTTCCA	AGTTGGGGT	TCTCCAAGGT	CATTGGGATT	2640
GCTTTCACAT	CCATTGCTA	TGTACCTTCC	CTATGATGGC	TGGGAGTGGT	CAACATCAA	2700
ACTAGGAAAG	CTACTGCCCA	AGGATGTCC	TACCTCTATT	CTGAAATGTG	CAATAAGTGT	2760
GATTAAGAG	ATTGCCTGTT	CTACCTATCC	ACACTCTCGC	TTTCAACTGT	AACTTTCTTT	2820
TTTTCTTTT	TTCTTTTTT	CTTTTTTTT	GAAACGGAGT	CTCGCTCTGT	CGCCCAAGCT	2880
AGAGTGCAGT	GGCACGATCT	CAGCTCACTG	CAAGCTCTG	CTCCCGGGT	CACGCCATT	2940
TCCTGCTCTA	CCCTCCCAAG	CAGCTGGGAC	TACAGGGGCC	TGCCACCATG	CCCAGCTAAT	3000
TTTTGTATT	TTTAGTAGAG	ACGGGGTTTC	ACCGCTGTTAG	CCAGGATGGT	CTCGATCTCC	3060
TGAACCTGTG	ATCCGCCCGC	CTCAGCCTCC	CAAAGTGTG	GGATTACAGG	CGTGAGGCCAT	3120
CGCACCCGGC	TCAACTGTAA	CTTCTATAC	TGGTTCATCT	TCCCCTGTAA	TGTTACTAGA	3180
GCTTTGAAG	TTTGGCTAT	GGATTATTC	TCATTTATAC	ATTAGATTT	AGATTAGTTC	3240
CAAATTGATG	CCCACAGCTT	AGGGTCTCTT	CCTAAATTGT	ATATTGTAGA	CAGCTGCAGA	3300
AGTGGGTGCC	AATAGGGAA	CTAGTTATA	CTTTCATCAA	CTTAGGACCC	ACACTTGTG	3360
ATAAAGAACAA	AAGGTCAAGA	GTATGACTA	CTGATTCCAC	AACTGATTGA	GAAGTTGGAG	3420
ATAACCCCGT	GACCTCTGCC	ATCCAGAGTC	TTTCAGGCAT	CTTTGAAGGA	TGAAGAAATG	3480
CTATTTAAT	TTTGGAGGT	TCTCTATCAG	TGCTTAGGAT	CATGGGAATC	TGTGCTGCCA	3540
TGAGGCCAAA	ATTAAGTCCA	AAACATCTAC	TGGTTCCAGG	ATTAACATGG	AAGAACCTTA	3600
GGTGGTCCCC	ACATGTTCTG	ATCCATCCTG	CAAATAGAC	ATGCTGCAC	AACAGGAAAA	3660

GTGCAGGCAG CACTACCACT TGGATAACCT GCAAGATTAT AGTTCAAGT AATCTAACCA 3720  
 TTTCTCACAA GGCCCTATTG TGTGACTGAA ACATACAAGA ATCTGCATT GGCCTCTAA 3780  
 GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACTG 3840  
 GAGTGCTTGG CAGGGAGAC AGAGTCAGG ACTGCCAATC GAGCCAATAC AGCAGGCTTA 3900  
 CACAGGAACC CAGGGCCTAG CCCTACAAAC ATTATTGGGT CTATTCAGTG TAAGTTTAA 3960  
 TTTCAGGCTC CACTGAAAAGA GTAAGCTAAG ATTCTGGCA CTTCTGTCT CTCTCACAGT 4020  
 TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080  
 CTTTGGGAGG CGGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140  
 ACAAAAGTGGAG ATACCCCCCTG ACCCCTTCTC TACAAAAATA AATTTTAAA ATTAGCCAA 4200  
 TGTGGTGGTG TATACTTACA GTCCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260  
 AGCCCAGGAA TTCAAGGCTG CAGTGAGCTA TGATTTCAAC ACTGCACCTC TGGCTGGCA 4320  
 ACAGAGCGAG ACCCTGTCTC AAAGCAAAAA GAAAAAGAAA CTAGAACTAG CCTAAGTTG 4380  
 TGGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440  
 ATTAGCCAA AAAGCTTGTG GTCTTGCTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500  
 CACCACTCAA TGGGAGGAGA GAGAAGTAAG CTGTTTGATG TATAGGGAA AACTAGAGGC 4560  
 CTGGAACATGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTGGAGTT AAGAAGGAGA 4620  
 GGAGGTCAGT ACTGCTGTT AGAGATTTT TTTATGTAAC TCTTGAGAAG CAAAACTACT 4680  
 TTTGTTCTGT TTGGTAATAT ACTTCAAAAC AAACCTCATA TATTCAAATT GTTCATGTCC 4740  
 TGAAATAATT AGGTAATGTT TTTTCTCTA TAG 4773

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8835 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: intron
- (B) LOCATION: 1..8835
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GTAAGAAAATA TCATTCCCTCT TTATTTGGAA AGTCAGCCAT GGCAATTAGA GGTAATAAG 60  
 CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT 120  
 CAACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTTT TCTTTTATCA GATATGGGAG 180  
 AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTTGCTCT CATGTTCCCT TTACACTAAG 240  
 CACATCACAT GTTAGCATAA GTAAACATTT TAATTAAGAA TAATATGTGTA CTTTTTAAC 300  
 AACAAAAAAA AGCATAAAGA GTGACACTTT TTTATTTTA CAAGTGTTTT AACTGTTTA 360  
 ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTTG CATTCAACT ACTGCAATAT 420  
 TGACAGAAAT GCAGCCTCTG GTAAACTCTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA 480  
 GACAAATTAC GTCTTAGAAT TATTAGAAAT AGCTTTCACT TTAGGAACTC CCTGAGAATT 540  
 GCTGCTTAG AGTGGTAAGA TAAATAAGCT TCTCTTAAA CGGAATCTCA AGACAGAAC 600  
 AGTTACATTA AAAGCAAACA AAAAATTGC CCATGGTTAG TCATCTGTG AAATCTGCCA 660  
 CACCTTGGA CTGGGCTACA ATTGGATAAT ATAGCATTCC CCGAGATAAT TTTCTCTCAC 720  
 AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC 780  
 CCCTAAATT TAGGGCTCT GAAATTGTC TTTTGCTTA TATTCACTA CTTTACGTT 840  
 TATTAATCT TCTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCC 900  
 AGCCCAGGAA TTGAGACCA GCCAGGGCAA CACAGTCTCT ACAAAAAAAAT AAAAATTAC 960  
 CTGGGTGTGT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGGCCA 1020  
 GGATAGCCAA ATCTGTGGTG AGTTCAGCCA CTAAACAGAG CGAGACTTTC TCAAAAAAAAC 1080  
 AAACAAAAAA ACAAAACAAAC TTCCCTCAAA ATAACTTTT ATCTGCAATG TTTTCCATT 1140  
 GCCTGTGAGA TAAATTTAC TCTTTTACCT GATTTCAAA GCCCTCCATA ATCTAATCCG 1200  
 ACTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260  
 GGTTGGGTGC AGTGGCTCAC TCCTGTAATC CCAACACTTT GGAAGGCCAA GGCAGGTGGA 1320  
 TTGCTTCAGC TCAGGAGTTC AAGACCAGCC TGGGCAACAT GGCAAAACC CTGTCTCTCC 1380  
 AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCCT GTAGTCCCAA CTACTCAAA 1440  
 GGCTAAGGCA AGAGGATCAC TTGAGCCAG GAGGTCAAGG CTACAGTGAG CCATGTTAC 1500  
 TGTGTCAGT CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAA AAAAAAAA 1560

GAAAAAGAAAAA GAAAAAAAACA TCGCTCTATT CAGTCACCC CCACCACAAC ATTGTTTGA 1620  
 TTATCACATA AATGCTGGTC CATTGCCCTC TCTATCTATT CAAATCTTA AGCATTCTT 1680  
 GAGATTCAAC TCAATTCTCC TTTTCAAAC AGGCCATTAA AACTACATCA GTTCCATT 1740  
 GATTTTCTTG CTTTGAGTCT ACAGACTCAA AAACAAAAAC TTAAAAACTT ATTTTTAAG 1800  
 TTTTCTGCTA CTCTCACTTC TTCACACCTC ACATACACGC ATTACATAATA AGATGGCAGA 1860  
 ATGTTCAAGG ATAAAATGAT TTATAGAACT GAAAAGTTAG GTTTGATCT TGTTGCTGTC 1920  
 AAGATGACTA CCTACCTGAT CTCAGGTAAT TAATTATGTA GCATGCTCCC TCATTCATC 1980  
 CCATACCTAT TCAACAGGAT TGGAATTCCA CAGCAAGGAT AAACATAATC ATAGTTGCTT 2040  
 TTCAAGTTCA AGGCATTTA ACTTTTAATC TAGTAGTATG TTTGTTGTTG TTGTTGTTGT 2100  
 TTGAGATGGA GCCCTGCTGT GTCACCCAGG CTGGAGTGC GTGGCACGAA CTCGGCTCAC 2160  
 TGCAACCTCT GCCTCATGGG TTCAATCAGT TATTCTGCCT CAGTGTCCC AGTAGCTGGG 2220  
 ACTACAAGGC ACATGCCACC ATGCCCTGGCT AATTTTGTAA TTTTAGTAG AAACAGGGCT 2280  
 TCACCAGTGT GGCCAGGCTG GTCTGAACT CCTGACCTCA AGTGATCCAG CGCGCTCGGC 2340  
 CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGC CAGCCTATAA GTATGTTTT 2400  
 AACTCTTAG TGGCTTAACA ATGCTGGTTG TATAATAAT ATGCCATAAA TATTTACTGT 2460  
 CTTAGAATTG TGAAGAAGTG GTTACTAGGC CGTTTGCCAC ATATCAATGG TTCTCTCCTT 2520  
 ACAGCTTAA TTAGAGTCTA GAATTGCAGG TTGGTAGAGC TGGAACAGAC CTTAAAGATT 2580  
 GACTAGCCAA CTTCTTGTC CAAATGAGGG AACTGAGACC CTTAAAATTA AGTGACTGTC 2640  
 CCCAGACAAA ACTGGAACTC ATGTGTCCTA ATTTCCATCA TGAAATTCTA CCATTCACTA 2700  
 GCCTCTGGCT AGTTGTCAAA GTATTGCATA ACTAAATT TATGCTGTT TTAAAGAAC 2760  
 AATTGTAAC TCTTACTCCT GGGAGGGTCT TTCTGAGGTG GTTTATAACT CTTAAAAAAA 2820  
 AAAAAGTCAG TAGTCTGAGA ATTTTAGACG AAATAGTCAA AGCATTTTA TCCAATGGAT 2880  
 CTATAATTTC CATAGATTAG AGTTAAATCA AAGAAACACG GATGAGAACAGA 2940  
 ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACTACTG TAATCAGTCA TAGATGTACT 3000  
 GAGAACTAAC AAGAAGAATT GTAAGAAATT AAGAATGAAG AATTCAAAAT CAACACATGA 3060  
 AATAAAAAGA AACTACTAGG GAAAATGGA GAAGACATTA GAAAATTAT TCTATTTTA 3120  
 AAATTCTGTT TTCAGGCTTC CCTCCTGTT TCATGGTTT TCAGGTGGAG 3180  
 GGAAAGTTA AGATGGAAAA AATATATATA TTCTACACAT CCCTTCTAC GCTGTTGTCA 3240  
 TGGCAACAAAG GTTATCATA GCAAACCTTT ATTACACAA CATTATTGA GTTCTTACTG 3300  
 TGTGGTAAGC TCTTCCAGG TGTTGAAAAT TCAGGGGAAA AAAGACAAC CATTGTCTTA 3360  
 AAACTCAGAT GAAAGCTGAA CAGACCTATT TTAATCAAA GTATCTCAA TTTAGGGTAG 3420  
 TAAGAGCTAT TTAAGAAGCA TGAACAGGTG TGAAGGAGGT AGGACTCTGA GGAGAGAATA 3480  
 GTTAGCTAGG AATGAAAGAG CAGAGAAGTT TCCCTAGAGG AACTATTAAA GCTGGGAGTT 3540  
 ACGGGATGAA AGATGAGGCA GGGTTGCG CAAAAAAA AAAAAAGGCA GGGGAAGGGG 3600  
 AAGTTCTGGC CTGGCAGAGA GAATACTGT GGCACAAATG GAGGAGAGTC TGGAAGCAAG 3660  
 AAAACCAAGT AGAAGAGTAT TAAAATAGAA GATGCCAGGG GTAATGAGGG CTTGATTAA 3720  
 AACAGTGTG TGAGGAGATGG AGAGGAGATA CAAATTCTG GAGACATTG TGAGTTAGAA 3780  
 CCTACAGTAT TTATCAGACAA AGGGAAAGAT TAGACAAAGG AGTTAAGAAT GACTCCCAGG 3840  
 TTTCAGTTG GGGCAGGTAA CTAGGACATG TTTGAAAAG TAATGTATTG GATCTCTTAC 3900  
 CATTGGAACAT ATGTATGTGG AGCCAAATTAA AAATTTGTAC ATGTATATAA CTCTCCCCC 3960  
 ACCACCAAGTA ACTACTTCCC TAACTCTCTA CTTGTAGCC AGACTTCTTA AAAGAATAGT 4020  
 TTGTAGTCAC TGTCTTTACT TTTCCCTCTC CATTCTGTCC TAGATATTG TCCACCTACC 4080  
 ATCTGCTGCC TCCACTTTAC CCAAACGTG CTACGGTGC CCAAACACTTC CTAATTGCCA 4140  
 AATTCAATGA ACAAGTTAA GCTTATATGT AAATTAGGAG CTCTACAGTT TGATTCGAG 4200  
 CAGCCCCCTCC TGAAACCTCT TCTCTTCA CTTCTGTGAC ACATCTCAGA TTTACAAAAC 4260  
 TGAACTAATT ATTTTACACT TGAGCTGTAT TTCTGTTCTT CTTCTTGTAT GAATGAGGTA 4320  
 ACCACTCAAC AAATTGCCCA AGCCAAAAC TACGAAGTCA TCCTCAGTTC CTCCTCTTC 4380  
 TGTTTGACCC ACAACAGATC AGCTGAGAAA TCCCGCTGTT TAGTATCTCT TGAATTCTT 4440  
 ACCTTAATTG ATAGCCTCAT CAACTCTTAA TTGTTAAAAT TACTTCAGTA GTTGTGTTCT 4500  
 GACCTCTGTC CAATCTTGTG CAATCAGGTC CATTCTTTG TTCTTGGTGG TGGTGGTGG 4560  
 GTTGACAGAG TTTCGCTTTT GCTGCCAGG CTGAAGTCA GTGGAGCACT TCACTGCAAC 4620  
 CACAGCCTCC TGGTTTAAG CAGTCACCC TCCCGAGTAG CTGGGACTAC AGGTATGTGC 4680  
 CACCACACCC AGCTAATTG GTGTTTCAG TAGAGACAGG GTTCACCAT GTTGGTCAGG 4740  
 CTGGTCTCAA ACTCCTGACC TCAAGCAATC CACCCACCTC AGCCCTCCAA AGTGCTGGGA 4800  
 TTACAGGCAT GAGCCACTGC ACACGGACCA GATCCATTGT TTATGTTGCT TCTAGAGTGA 4860  
 GTTTTAAAAA CACAAATTG ACCATATCTT TCTCCATT TAAAGTCAGTAT TTTTTTTTC 4920  
 AGGAAAAAAC AGTCAAACT CTTTAGTCTG CTTACACAAAG GCCTTGTAG TCTGACTCTT 4980  
 CTTTCCAAGC TTTCATCAA GTATACTGCA AGTTACATT TATGTTGAATT GAATTAGGCA 5040  
 ACGGTATAAA AATTATAGTT TATATGGGCA AAATGGAAAT AATGTTAACT CTTCCAAATA 5100  
 GTTTATCTAG AATGACATAA TTCAAGACT GTCAGGTCAA ATGAGTTATA AACTGTTAAC 5160  
 ACTATTGCCA CATGCAAGTG TCTCTTATAC TTGGTAAAT TATCTGCTTC CATGTCATTA 5220  
 TTATGTAAT TAGACTTTAA ATAACCTCAGA AGTTCTTCAG ACATACAGGT TATTATTGTG 5280  
 CTTTTAAAC ATAATTAA ATAATTAAAT ATATGATAAT GTTATCCAAG TGCTAAGGGA 5340  
 TGTATTGTTA CTGCTGTGCA AAAAAAAA AAAAAAAAC TCCAAATAAA TATGTTGAA 5400  
 CCAAGTTAT ATGCAAGAAA ACAATATTAA AAAGGCCAAA GTACCACCAT AATAGGCTGT 5460  
 GTGGAGACGG CAGGCTACAA AACACTAGTA ATAATGCTGA GAAAGTTGAA AAAAGAAAGA 5520  
 AAGCAACAAAT ATGCTTTGGT TGTTGTAGGT TTATGTTACTC CAAGAATATC TCCTCTCAA 5580

CTTTTACGTT TTTCCAAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640  
 GTCGGCAGCCT TTGGGAGGCC AAGGCGGCA GATCACCTGA GGTCAAGGAGT TTGAGACCG 5700  
 CCTGACCAAA ATGGAGAAA CCCGCCCCC TCAACTACTAA AAGAATACAA AATTAGGCC 5760  
 GGCACTGG CTTACCCCTG TGATCCCAGC ACTTTGGAG GCCGAAGCAG GAAGATCACC 5820  
 TGAGGTCAAG AGTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAA ATACAAAATT 5880  
 AGCCGGCGT GGTGGTCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940  
 CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTGCCATTG CACTCCAGCC 6000  
 TGGGCAACAA GAGCGAAACT CTGTATCCAA AAAACAAAAG AAAAGAAAAG GTAACCTTGA 6060  
 ACTATGTGAG ATCTTAGAA ATGCATCTT TCTGTAAAAT GTGACTACAT TTGCCTTATT 6120  
 TATGGTAAAA ATGGTGAGGC CTCAAACAAAC CCATATTTTC TCAGGTCTCCC CGCTGCCTAG 6180  
 CCTTGTCA CATTGCTTCT TCTTGGTGA AGCTCTTCCT CTGGCCTTGA AAATGCCTGC 6240  
 TTCTCTTCA AGGTAGCACA GTCATCACT TCTGTGGTAA CCTTCTCCAG CACCATCAAA 6300  
 CAGAAAGAAT GAATCTCTTG TAAATTTCAGC TCTTACGTCA TTCATTACAT TATTTGTAA 6360  
 CTCTTATAG ATTCTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACTC 6420  
 TCATTCACTG GTGGTTGGT CAGCTACTGG CCACATTCT GATGCTAGT TAATGCTCAA 6480  
 ACCTTAACG GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCATTCA CTCATTGACT 6540  
 AACTATGTAC TCAAATAGT AACACCCAGT AATTAAATCC AATTCCCTGCC CATACTGCTT 6600  
 GGTACATTC AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660  
 TACAGAAGAT CATGCTAATC ATAATTCAAC ACTGATAACT AATCAAACAT AAATGCTCTC 6720  
 AGGTTAACAA ATGTCTGCCT TCTCAGTTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780  
 GATAATAGGG CCTTGTTCAG CAATGAAGCC ATAAAGGTGA ATAAAGAACAA TGCCCTCGTG 6840  
 GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTTAATAT TAATGTTGGT TTGGGTTTG 6900  
 GTGAAAATG TTAGACTTA TCTTAGTGTAT CTTTCATCC TTTGCTATAT TATTTTCTC 6960  
 TAAGAGTCTT CCTTATCCCC TCTTTAAAAA AACTAGGTGA TAATTCTAAA TTGTAATTTT 7020  
 AAAATTATA AATAGCTTAT AAAATTAAAT ATTATAATAA TTAAATGTT TGATAAATAT 7080  
 TTAAATTTA TAATATTAA ATGTTTATTAA AATTCAATT GTACATCACT TTTTATTAA 7140  
 TTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCCAGAACT TTGAGAGGCC 7200  
 AAGTCAGGCA AACCATTGA GCTCAGGAGT TTGAGACCCAC CCTGGGCAAC GTGGTGAAC 7260  
 CCTGTCTCTA CCAAACATAT GAAAACATTAT CTGGGTGTGG TGGCACGCAT CTGTGGCTCC 7320  
 AGATGGGAGT CCCAGGCTAA GATGGGAGAA TCGCTTGAAC CCAGGTGAGA GGGGTGGGT 7380  
 GGATGTTGCA GTGAGCTGAG ATCGTGCAC TGCACCTCAA CCTGGGTGAC AGAGTGAGAC 7440  
 TCCATCTCAA AAAAAAAATG TGTTATCTAA ATAAGATAAA TTAATAACT GTTCGCACTT 7500  
 AGATGAGCAT AAGGAACCTAA ACCTAGATAA AACTATCAA TAAGGCCTGG GTACAGTGAC 7560  
 TCATGCCCTGT AATCTCAAGC ACTTTGGAG GCCAAATTTA TACAAAGTT GTTGTATAAC 7620  
 ACCAACTAAC AACTATTTG GGGTTAGCTT AATTCAAGATT AATTTTTTTAA AACTGAGTT 7680  
 TTAAATTCTC GCTTACTCTA CCATACATGC TAGGCCTCAT ATTATGCTAG AAAAATTTC 7740  
 AGCACAGATT TATGAATACT CTCCTGCATA CCATTAAATT TTTAAACAAA TTTAAATGCA 7800  
 GTATATATGT GCCTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860  
 TCTGTAATT CAAAGTAGTA ATGAGTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920  
 AGCACAACAG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980  
 TGTGGTTTGT TTCCCTGCAAC TAATCATGGG AGGAATGCTA AATTTCAGAG GTTGGTGAAC 8040  
 ATACATGTGT ATTTTTTCTC CCATCCAAGT TCACAGATT CTCACACTGA GAACTCCTAT 8100  
 TCCATAACAA AATTCTGGAA GCCTGCACAC CGTATTGGAA GAAAGGGCAGA AAGGAAAAGC 8160  
 AAATGGAAGG ATTTAAATT TTTCAAATC CTGTATCCCT TGATTTACA GCAAGATTGT 8220  
 ATTTATGTAT TACTGTGT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280  
 GCAGCTCAGG GAGAAAGAGG GCCACCAAA GCCAGAGCCC TTCAGCTTC TCCCACCCCTG 8340  
 CCTGTACCCCT CAGATGGAAAG CACTTTTA TCATTGTTTC ACCTTTAGCA TTTTGACAAT 8400  
 GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460  
 AGCCAGCTCT TCTAAAGGG CACGATTAGA TGTGTTTATG GCATCCTCAG GTGAAACTAT 8520  
 ATTTATATTG ACAATATATT TATATTCTC AAGGAATACT AGAATAATGA TTCAGTTCA 8580  
 TACTAGGCCA TTATCTTAC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640  
 CAAATATCTG ATGATTTGTA AGAGAACACT TAAACATGGG TATTCTATAAG CTGAAACTTC 8700  
 TGGCATTAT TGATGTCAA GATTGTTCAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760  
 ACTTGAAGGT AGTATAAAAGT AGTACTAAAA GGTACAATCA TTGTCCTTA ACAGATGGCT 8820  
 CTTTGCTTTC ATTAG 8835

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1371 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human  
(F) TISSUE TYPE: placenta

(ix) FEATURE:  
(A) NAME/KEY: intron  
(B) LOCATION: 1..1371  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTAAGGCTAA	TGCCATAGAA	CAAATACCAAG	GTTCAGATAA	ATCTATTCAA	TTAGAAAAGA	60
TGTTGTGAGG	TGAACATATTA	AGTGACTCTT	TGTGTCACCA	AATTCACTG	TAATATTAAT	120
GGCTCTTAAA	AAAATAGTGG	ACCTCTAGAA	ATTAACCACA	ACATGTCCAA	GGTCTCAGCA	180
CCTTGTACCA	CCACGTGTCC	TGGCACTTTA	ATCAGCAGTA	GCTCACTCTC	CAGTTGGCAG	240
TAAGTGCACA	TCATGAAAAT	CCCAGTTTC	ATGGGAAAAT	CCCAGTTTC	ATTGGATTTC	300
CATGGGAAAAT	ATCCCACTAC	AAAACGGGT	GCATTCAAGGA	AATACAATT	CCCAAAGCAA	360
ATTGGCAAT	TATGTAAGAG	ATTCTCTAAA	TTTAGAGTTC	CGTGAATTAC	ACCATTTAT	420
GTAAATATGT	TTGACAAGTA	AAAATTGATT	CTTTTTTTT	TTTCTGTTG	CCCAGGCTGG	480
AGTGCAGTGG	CACAACTCTC	GCTCACTGCA	ACCTCCACCT	CCTGGGTTCA	AGCAATTCTC	540
CTGCCTCAGC	CTTCTGAGTA	GCTGGACTA	CAGGTGCATC	CCGCCATGCC	TGGCTAATT	600
TTGGGTATTT	TTACTAGAGA	CAGGGTTTG	GCATGTTGTC	CAGGCTGGTC	TTGGACTCCT	660
GATCTCAGAT	GATCCTCCTG	GCTCGGGCTC	CCAAAGTGCT	GGGATTACAG	GCATGAACCA	720
CCACACATGG	CCTAAAAATT	GATTCTTATG	ATTAATCTCC	TGTGAACAAT	TTGGCTTCAT	780
TTGAAAGTTT	GCCTTCATT	GAAACCTTCA	TTTAAAAGCC	TGAGCAACAA	AGTGAGACCC	840
CATCTCTACA	AAAAACTGCA	AAATATCCTG	TGGACACCTC	CTACCTCTG	TGGAGGCTGA	900
AGCAGGAGGA	TCACCTGAGC	CTAGGAATT	GAGCCTGAG	TGAGCTATGA	TCCCACCCCT	960
ACACTCCAGC	CTGCATGACA	GTAGACCCCTG	ACACACACAC	ACAAAAAAA	ACCTTCATAA	1020
AAAATTATTA	GTTGACTTTT	CTTAGGTGAC	TTTCCGTTA	AGCAATAAT	TTAAAAGTAA	1080
AATCTCTAA	TTTAGAAAAT	TTATTTTAG	TTACATATTG	AAATTTTAA	ACCCTAGGTT	1140
TAAGTTTAT	GTCTAAATTA	CCTGAGAACCA	CACTAAGTCT	GATAAGCTTC	ATTTTATGGG	1200
CCTTTGGAT	GATTATATAA	TATTCTGATG	AAAGCCAAGA	CAGACCTTA	AACCATAAAA	1260
ATAGGAGTTC	GAGAAAGAGG	AGTAGCAAAA	GTAAAAGCTA	GAATGAGATT	GAATTCTGAG	1320
TCGAAATACA	AAATTTACA	TATTCTGTTT	CTCTCTTTT	CCCCCTCTTA	G	1371

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3383 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: human  
(F) TISSUE TYPE: placenta

(ix) FEATURE:  
(A) NAME/KEY: intron  
(B) LOCATION: 1..3383  
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTAAAGTAGA	AATGAATT	TTTTCTT	CAAACTAAGT	ATCTGTTGA	GACACATCTA	60
TCTCACCAT	GTCACTGAG	GA	AA	ATGGTTCTC	ATGCTACCA	120
AAGAAATGTG	GACTCAGTAG	CACAGCTT	GAATGAAGAT	GATCATAAGA	GATACAAAGA	180
AGAACCTCTA	GCAAAAGATG	CTTCTCTATG	CCTTAA	TTCTCCAGCT	CTTAGAATCT	240
ACAAAATAGA	CTTGCCTG	TTCATTGGTC	CTAAGATTAG	CATGAAGCCA	TGGATTCTGT	300
TGTAGGGGA	GCGTTGCATA	GGAAAAGGG	ATTGAAGCAT	TAGAATTGTC	AAAATCAGT	360
AACACCTCCT	CTCAGAAATG	CTTGGGAAG	AAGCCTGGAA	GGTCCGGGT	TGGTGGTGGG	420
GTGGGGCAGA	AAATTCTGGA	AGTAGAGGG	ATAGGAATGG	GTGGGGCAAG	AAGACCACAT	480
TCAGAGGCCA	AAAGCTGAAA	GAAACCATGG	CATTATGAT	GAATTCA	TAATTCA	540
TGGAAGTAGA	GTAGGAGTAG	GAGACTGGT	AGAGGAGCTA	GAGTGATAAA	CAGGGTGTAG	600
AGCAAGACGT	TCTCTCACCC	CAAGATGTGA	AATTGGACT	TTATCTTGG	GATAATAGGG	660
TTAATTAAGC	ACAATATGTA	TTAGCTAGGG	TAAAGATTAG	TTTGTGTAA	CAAAGACATC	720

CAAAGATACA	GTAGCTGAAT	AAGATAGAGA	ATTTTCTCT	CAAAGAAA GT	CTAAGTAGGC	780
AGCTCAGAAG	TAGTATGGCT	GGAAGCAACC	TGATGATATT	GGGACCCCCA	ACCTTCTTC	840
GTCTTG TAC	CATCATCCCC	TAGTTGTTGA	TCTCACTCAC	ATAGTTGAAA	ATCATCATAC	900
TTCCCTGGGTT	CATATCCCAG	TTATCAAGAA	AGGGTCAAGA	GAAGTCAGGC	TCATTCCTT	960
CAAAGACTCT	AATTGGAAGT	TAAACACATC	AATCCCCCTC	ATATTCCATT	GACTAGAATT	1020
TAATCACATG	GCCACACCAA	GTGCAAGGAA	ATCTGGAAA	TATAATCTT	ATTCCAGGTA	1080
GCCATATGAC	TCTTTAAAT	TCAGAAATAA	TATATTTTA	AAATATCATT	CTGGCTTGG	1140
TATAAAGAAT	TGATGGTGTG	GGGTGAGGAG	GCCAAAATTA	AGGGTTGAGA	GCCTATTATT	1200
TTAGTTATTA	CAAGAAATGA	TGGTGTATG	AATTAAGGT	GACATAGGGG	AGTGCTGATG	1260
AGGAGCTGTG	AATGGATTT	AGAAACACTT	GAGAGAACATCA	ATAGGACATG	ATTAGGGTT	1320
GGATTTGGAA	AGGAGAAGAA	AGTAGAAAAG	ATGATGCCTA	CATTTTCAC	TTAGGCAATT	1380
TGTACCATTC	AGTGAATAG	GGAACACAGG	AGGAAGAGCA	GGTTTGGTG	TATACAAAGA	1440
GGAGGATGGA	TGACGCATT	CGTTTGGAT	CTGAGATGTC	TGTGGAACGT	CCTAGTGGAG	1500
ATGTCACAA	ACTCTTCTAC	ATGTGGTTCT	GAGTTCAGGA	CACAGATTG	GGCTGGAGAT	1560
AGAGATATTG	TAGGCTTATA	CATAGAAATG	GCATTTGAAT	CTATAGAGAT	AAAAAGACAC	1620
ATCAGAGGAA	ATGTGTAAG	TGAGAGAGGA	AAAGCCAAGT	ACTGTGCTGG	GGGAATACC	1680
TACATTTAAA	GGATGCAGTA	GAAAGAAGCT	AATAAACAC	AGAGAGCAGA	CTAACCAAAA	1740
GGGGAGAGA	AAAACCAAGA	GAATTCCACC	GACTCCCAGG	AGAGCATTTC	AAGATTGAGG	1800
GGATAGGTGT	TGTGTTGAAT	TTTGAGCCT	TGAGAATCAA	GGGCCAGAAC	ACAGCTTTA	1860
GATTTAGCAA	CAAGGAGTTT	GGTGATCTCA	GTGAAAGCAG	CTTGATGGTG	AAATGGAGGC	1920
AGAGGCAGAT	TGCAATGAGT	GAAACAGTGA	ATGGGAAGTG	AAGAAATGAT	ACAGATAATT	1980
CTTGCTAAAA	GCTTGGCTGT	TAAAAGGAGG	AGAGAAACAA	GACTAGCTGC	AAAGTGAGAT	2040
TGGGTTGATG	GAGCAGTTT	AAATCTCAA	ATAAAGAGCT	TTGTGCTTTT	TTGATTATGA	2100
AAATAATGTG	TTAATTGTA	CTAATTGAGG	CAATGAAAAA	AGATAATAAT	ATGAAAGATA	2160
AAAATATAAA	AACCACCCAG	AAATAATGAT	AGCTTACATT	TTGATACAAT	ATTCTACAC	2220
TCCTTCTAT	GTATATATAC	AGACACAGAA	ATGCTTATAT	TTTTTATTAA	AGGGATTGTA	2280
CTATACCTAA	GCTGCTTTT	CTAGTTAGTG	ATATATATGG	ACATCTCTCC	ATGGCAACGA	2340
GTAATTGCAG	TTATATTAAG	TTCATGATAT	TTCACAAATAA	GGGCATATCT	TTGCCCTTT	2400
TATTTAATCA	ATTCTTAATT	GGTGAATGTT	TGTTTCCAGT	TTGTTGTTGT	TATTAACAAT	2460
GTTCCCATAA	GCATTCCCTGT	ACACCAATGT	TCACACATTT	GTCTGATTT	TTCTTCAGGA	2520
TAAAACCCAG	GAGGTAGAAT	TGCTGGTTG	ATAGAAGAGA	AAGGATGATT	GCCAAATTAA	2580
AGCTTCAGTA	GAGGGTACAT	GCCGAGCACA	AATGGGATCA	GCCCTAGATA	CCAGAAATGG	2640
CACTTTCTCA	TTTCCCCTTG	GGACAAAAGG	GAGAGAGGCA	ATAACTGTGC	TGCCAGAGTT	2700
AAATTGTCAC	GTGAGGTAGC	AGGAAATCAT	TTGCTGAAA	TGAAAACAGA	GATGATGTTG	2760
TAGAGGTCTC	GAAGAGAGCA	AAGAAAATT	GAATTGCGG	CTATCAGCTA	TGGAAGAGAG	2820
TGCTGAACGT	GAAAACAAA	GAAGTATTGA	CAATTGGTAT	GCTTGTAAATG	GCACCGATT	2880
GAACGCTGT	GCCATTGTT	ACCAGCAGCA	CTCAGCAGGC	AAGTTGGAG	TTTTGTAGCA	2940
GAAAGACAAA	TAAGTTAGGG	ATTAATATC	CTGGCCAAT	GGTAGACAAA	ATGAACCTCG	3000
AGATCCAGCT	GCACAGGGAA	GGAAGGGAAG	ACGGGAAGAG	GTTAGATAGG	AAATACAAGA	3060
GTCAGGAGAC	TGGAAGATGT	TGTGATATT	AAGAACACAT	AGAGTTGGAG	AAAAGTGTA	3120
AGAAAACATG	AAGGGTAAGA	GACCGGTCAG	AAAGTAGGCT	ATTTGAAGTT	AAACACTTCAG	3180
AGGCAGAGTA	GTTCTGAATG	GTAACAAGAA	ATTGAGTGTG	CCTTGAGAG	TAGGTTAAA	3240
AACAATAGGC	AACTTTATTG	TAGCTACTTC	TGGAACAGAA	GATTGTCATT	AATAGTTTA	3300
GAAAACCTAA	ATATATAGCA	TACTTATTG	TCAATTAAACA	AAGAAACTAT	GTATTTTAA	3360
ATGAGATTTA	ATGTTTATTG	TAG				3383

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11464 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
  
- (ii) MOLECULE TYPE: Genomic DNA
  
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: human
  - (F) TISSUE TYPE: placenta
  
- (ix) FEATURE:
  - (A) NAME/KEY: 5'UTR
  - (B) LOCATION: 1..3
  - (C) IDENTIFICATION METHODS: E
  - (A) NAME/KEY: leader peptide
  - (B) LOCATION: 4..82
  - (C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron  
 (B) LOCATION: 83..1453  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: leader peptide  
 (B) LOCATION: 1454..1465  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 1466..4848  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: leader peptide  
 (B) LOCATION: 4849..4865  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 4866..4983  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 4984..6317  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 6318..6451  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: intron  
 (B) LOCATION: 6452..11224  
 (C) IDENTIFICATION METHODS: E  
 (A) NAME/KEY: mat peptide  
 (B) LOCATION: 11225..11443  
 (C) IDENTIFICATION METHODS: S  
 (A) NAME/KEY: 3'UTR  
 (B) LOCATION: 11444..11464  
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA	48
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala	
-35 -30 -25	
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT	98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	
AGAACAAATA CCAGGTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAAC	158
ATTAAGTGAC TCTTGTGTC ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAAATA	218
GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT	278
GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA	338
AAATCCCAGT TTTCATGGGA AAATCCCAGT TTTCATTGGA TTTCCATGGG AAAATCCCA	398
GTACAAAACT GGGTGCATTC AGGAAATACA ATTCCCAA GCAAATTGGC AAATTATGTA	458
AGAGATTCTC TAAATTAGA GTTCCGTGAA TTACACCATT TTATGTAAT ATGTTGACA	518
AGTAAAAATT GATTCTTTT TTTCCTTCT GTGCCAGG CTGGAGTGCA GTGGCACAAAT	578
CTCTGCTCAC TGCAACCTCC ACCTCCGGG TTCAAGCAAT TCTCTGCTC CAGCCTCTG	638
AGTAGCTGGG ACTACAGGTG CATCCGCCA TGCCCTGGCTA ATTTCGGGT ATTTTACTA	698
GAGACAGGGT TTTGGCATGT TGTCCAGGCT GGTCTGGAC TCCTGTCTC AGATGATCCT	758
CCTGGCTCGG GCTCCCAAAG TGCTGGATT ACAGGCATGA ACCACCACAC ATGGCCTAAA	818
AATTGATTCT TATGATTAAT CTCCTGTGAA CAATTGGCT TCATTTGAAA GTTGCCTTC	878
ATTGAAACC TTCATTTAAA AGCCTGAGCA ACAAAAGTGAG ACCCCATCTC TACAAAAAAC	938
TGCAAAATAT CCTGTGGACA CCTCCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACTT	998
GAGCCTAGGA ATTTGAGCCT GCAGTGAGCT ATGATCCCAC CCCTACACTC CAGCCTGCAT	1058
GACAGTAGAC CCTGACACAC ACACACAAAA AAAACCTTC ATAAAAAAATT ATAGTTGAC	1118
TTTTCTTAGG TGACTTCCG TTTAAGCAAT AAATTAAAAA GTAAAATCTC TAATTTAGA	1178
AAATTTATT TTAGTTACAT ATTGAAATT TAAACCTTA GGTTAAGTT TTATGCTAA	1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTTA TGGGCCTTT GGATGATTAT	1298
ATAATATCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAATAGGA GTTCGAGAAA	1358
GAGGAGTAGC AAAAGTAAAA GCTAGAATGA GATTGAATC TGAGTCGAAA TACAAAATT	1418
TACATATTCT GTTCTCTCT TTTCCCCCT CTTAG CT GAA GAT GAT G GTAAA	1470
Ala Glu Asp Asp Glu	
-10	
GTAGAAATGA ATTATTTTTT CTTGAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA	1530
CCATTGTCAG CTGAGGAAAA AAAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA	1590
ATGTGGACTC AGTAGCACAG CTTTGAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC	1650

CTCTAGCAAA	AGATGTTCT	CTATGCCCTA	AAAAATTCTC	CAGCTCTTAG	AATCTACAAA	1710
ATAGACTTG	CCTGTTCAT	TGGCCTAAG	ATTAGCATGA	AGCCATGGAT	TCTGTTGTAG	1770
GGGGAGCGTT	GCATAGGAAA	AAGGGATGAA	AGCATTAGAA	TTGTCAAAAA	TCAGTAACAC	1830
CTCCTCTAG	AAATGCTTG	GGAAGAACCC	TGGAAGGTT	CGGGTTGGT	GTGGGGTGGG	1890
GCAGAAAATT	CTGGAAGTAG	AGGAGATAGG	AATGGGTGGG	GCAAGAAGAC	CACATTCAAGA	1950
GGCCAAAAGC	TGAAAGAAC	CATGGCATT	ATGATGAATT	CAGGGTAATT	CAGAATGGAA	2010
GTAGAGTAGG	AGTAGGAGAC	TGGTGGAGG	AGCTAGAGTG	ATAAACAGGG	TGTAGAGCAA	2070
GACGTTCT	CACCCCAAGA	TGTGAAATT	GGACTTTATC	TTGGAGATAA	TAGGGTTAAT	2130
TAAGCACAAT	ATGTATTAGC	TAGGGTAAAG	ATTAGTTGT	TGTAACAAAG	ACATCCAAAG	2190
ATACAGTAGC	TGAATAAGAT	AGAGAATT	TCTCTAAAG	AAAGTCTAAG	TAGGCAGCTC	2250
AGAAGTAGTA	TGGCTGGAAG	CAACCTGATG	ATATTGGGAC	CCCCAACCTT	CTTCAGTCTT	2310
GTACCCATCA	TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	TGAAAATCAT	CATACTTCCT	2370
GGGTTCATAT	CCCAAGTTATC	AAGAAAGGGT	CAAGAGAAGT	CAGGCTCATT	CCTTTCAAAG	2430
ACTCTAATG	GAAGTTAAC	ACATCAATCC	CCCTCATATT	CCATTGACTA	GAATTTAAC	2490
ACATGGCCAC	ACCAAGTGCA	AGGAAATCTG	GAAAATATAA	TCTTATTCC	AGGTAGCCAT	2550
ATGACTCTT	AAAATTAGA	AATAATAT	TTTAAATAA	TCATTCTGGC	TTGGTATAA	2610
AGAATTGATG	GTGTGGGTG	AGGAGGCCAA	ATTAAAGGGT	TGAGAGCCTA	TTATTTAGT	2670
TATTACAAGA	AATGATGGTG	TCATGAATTA	AGGTAGACAT	AGGGGAGTGC	TGATGAGGAG	2730
CTGTGAATGG	ATTTAGAAA	CACTTGAGAG	AATCAATAGG	ACATGATT	GGGTTGGATT	2790
TGGAAAGGAG	AAGAAAGTAG	AAAAGATGAT	GCCTACATT	TTCACTTAGG	CAATTGTAC	2850
CATTCACTGA	AATAGGGAA	ACAGGAGGAA	GAGCAGGTT	TGGTGTATAC	AAAGAGGAGG	2910
ATGGATGACG	CATTCTGTT	TGGATCTGAG	ATGTCTGTG	AACGTCCTAG	TGGAGATGTC	2970
CACAAACTCT	TCTACATGTG	GTTCTGAGTT	CAGGACACAG	ATTGGGCTG	GAGATAGAGA	3030
TATTGTAGGC	TTATACATAG	AAATGGCATT	TGAATCTATA	GAGATAAAA	GACACATCAG	3090
AGGAAATGTG	TAAAGTGAGA	GAGGAAAAGC	CAAGTACTGT	GCTGGGGGGA	ATACCTACAT	3150
TTAAAGGATG	CACTAGAAAG	AAGCTAATAA	ACAACAGAGA	GCAGACTAAC	CAAAGGGGA	3210
GAAGAAAAAA	CAAGAGAATT	CCACCGACTC	CCAGGAGAGC	ATTTCAGAT	TGAGGGGATA	3270
GGTGTGTGT	TGAATTTCG	AGCCTTGAGA	ATCAAGGGCC	AGAACACAGC	TTTTAGATT	3330
AGCAACAAAG	AGTTGGTGA	TCTCAGTGA	AGCAGCTTG	TGGTGAATG	GAGGCAGAGG	3390
CAGATTGCAA	TGAGTGAAC	AGTGAATGGG	AAAGTGAAGAA	ATGATACAGA	TAATTCTTC	3450
TAAAAGCTG	GCTGTTAAA	GGAGGGAGAGA	AACAAGACTA	GCTGCAAAGT	GAGATTGGGT	3510
TGATGGAGCA	GTTTAAATC	TCAAAATAAA	GAGCTTGTG	CTTTTTGAT	TATGAAAATA	3570
ATGTGTTAAT	TGTAACTAAT	TGAGGCAATG	AAAAAAGATA	ATAATATGAA	AGATAAAAAT	3630
ATAAAAACCA	CCCAGAAATA	ATGATAGCTA	CCATTGAT	ACAATATTC	TACACTCCTT	3690
TCTATGTATA	TATACAGACA	CAGAAATGCT	TATATTTTA	TTAAAAGGGA	TTGTACTATA	3750
CCTAAGCTGC	TTTTCTAGT	TAGTGTATTA	TATGGACATC	TCTCCATGGC	AACGAGTAAT	3810
TGCAGTTATA	TTAAGTTCAT	GATATTTCAC	AATAAGGGCA	TATCTTGCC	CTTTTTATT	3870
AATCAATTCT	TAATTGGTGA	ATGTTGTTT	CCAGTTGTT	GTTGTTATT	ACAATGTCC	3930
CATAAGCATT	CCTGTACACC	AATGTTACA	CATTGCTG	ATTTTTCTT	CAGGATAAAA	3990
CCCAGGAGGT	AGAATTGCTG	GGTTGATAGA	AGAGAAAGGA	TGATTGCCA	ATTAAAGCTT	4050
CAGTAGAGGG	TACATGCCGA	GCACAAATGG	GATCAGCCCT	AGATACCAGA	AATGGCACTT	4110
TCTCATTCC	CCTTGGGACA	AAAGGGAGAG	AGGCAATAAC	TGTGCTGCC	GAGTTAAATT	4170
TGTACGTGGA	GTAGCAGGAA	ATCATTGCT	GAAAATGAAA	ACAGAGATGA	TGTTGTAGAG	4230
GTCCTGAAGA	GAGCAAAGAA	AATTGAAAT	TGCGGCTATC	AGCTATGGAA	GAGAGTGTG	4290
AACTGGAAAA	CAAAGAGAAGT	ATTGACATT	GGTATGCTTG	TAATGGCACC	GATTGAAACG	4350
CTTGTGCCAT	TGTCACCAG	CAGCACTCAG	CAGCCAAGTT	TGGAGTTTG	TAGCAGAAAG	4410
ACAAATAAGT	TAGGGATT	ATATCTGGC	CAAATGGTAG	ACAAAATGAA	CTCTGAGATC	4470
CAGCTGCACA	GGGAAGGAAAG	GGAGACGGG	AAGAGGTTAG	ATAGGAAATA	CAAGAGTCAG	4530
GAGACTGGAA	GATGTTGTG	TATTTAAGAA	CACATAGAGT	TGGAGTTAA	GTGTAAGAAA	4590
ACTAGAAGGG	TAAGAGACCG	GTCAGAAAGT	AGGCTATTG	AAGTTAACAC	TTCAGAGGCA	4650
GAGTAGTTCT	GAATGGTAAC	AAGAAATTGA	GTGTGCCTT	GAGAGTAGGT	AAAAAACAA	4710
TAGGCAACCT	TATTGTAGCT	ACTTCTGGAA	CAGAAGATTG	TCATTAATAG	TTTTAGAAAA	4770
CTAAAATATA	TAGCATACTT	ATTTGTCAAT	TAACAAAGAA	ACTATGTATT	TTTAAATGAG	4830
ATTTAATGTT	TATTGTAG	AA AAC CTG GAA TCA GAT	TAC TTT GGC AAG CTT			4880
		Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu				
	-5	1	5			
GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC						4928
Glu Ser Lys Leu Ser Val Ile Arg Asn Asp Gln Val Leu Phe						
10	15	20				
ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC						4976
Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp						
25	30	35				
TGT AGA G	GTATTTTT	TTAATCGCA	AACATAGAAA	TGACTAGCTA	CTTCTCCCCA	5032
Cys Arg Asp	40					
TTCTGTTTTA	CTGCTTACAT	TGTTCCGTGC	TAGTCCCAAT	CCTCAGATGA	AAAGTCACAG	5092
GAGTGACAAT	AATTCACCT	ACAGGAAACT	TTATAAGGCA	TCCACGTTT	TTAGTTGGGG	5152

TAAAAAATTG	GATACAATAA	GACATTGCTA	GGGGTCATGC	CTCTCTGAGC	CTGCCTTGA	5212
ATCACCAATC	CCTTTATTGT	GATTGCATTA	ACTGTTAAA	ACCTCTATAG	TTGGATGCTT	5272
AATCCCTGCT	TGTTACAGCT	GAAAATGCTG	ATAGTTTAC	AGGTGTGGTG	GCATCTATCT	5332
GTAATCCTAG	CTACTTGGGA	GGCTCAAGCA	GGAGGATTGC	TTGAGGCCAG	GACTTTGAGG	5392
CTGTAGTACA	CTGTGATCGT	ACCTGTGAAT	AGCCACTGCA	CTCCAGCCTG	GGTGATATAC	5452
AGACCTTGTG	TCTAAAATTA	AAAAAAAAAA	AAAAAAAAAC	CTTAGGAAAG	GAAATTGATC	5512
AAGTCTACTG	TGCCTTCAA	AACATGAATT	CCAATATCA	AAGTTAGGCT	GAGTTGAAGC	5572
AGTGAATGTG	CATTCTTAA	AAATACTGAA	TACTTACCTT	AACATATATT	TTAAATATT	5632
TATTTAGCAT	TTAAAAGTTA	AAAACAATCT	TTTAGAATT	ATATCTTAA	AATACTCAA	5692
AAAGTTGCAG	CGTGTGTGTT	GTAATACACA	TTAAACTGTG	GGGTTGTTG	TTTGTGAG	5752
ATGCAGTTTC	ACTCTGTAC	CCAGGCTGAA	GTGCAGTGC	GTGCAGTGGT	GTGATCTCGG	5812
CTCACTACAA	CCTCCACCTC	CCACGTCAA	GCAGATTCTCA	TGCCTCAGTC	TCCCAGTAG	5872
GTGGGATTAC	AGGCATGCAC	CACTTACACC	CGGCTAATT	TTGTATTTT	AGTAGAGCTG	5932
GGGTTTACCC	ATGTTGGCCA	GGCTGGCTC	AAACCCCTAA	CCTCAAGTGA	TCTGCCTGCC	5992
TCAGCCTCCC	AAACAAACAA	ACAACCCAC	AGTTTAATAT	GTGTTACAAC	ACACATGCTG	6052
CAACTTTAT	GAGTATTTA	ATGATATAGA	TTATAAAAGG	TTGTTTTAA	CTTTAAATG	6112
CTGGGATTAC	AGGCATGAGC	CACTGTGCCA	GGCCTGAAC	GTGTTTTAA	AAATGTCTGA	6172
CCAGCTGTAC	ATAGTCTCCT	GCAGACTGGC	CAAGTCTCAA	AGTGGGAACA	GGTGTATTAA	6232
GGACTATCCT	TTGGTTAAAT	TTCCGCAAAT	GTTCCTGTGC	AAGAATTCTT	CTAACTAGAG	6292
TTCTCATTTA	TTATATTTAT	TTCAG	AT AAT GCA CCC CGG ACC ATA TTT ATT			6343
		Asp Asn Ala Pro Arg Thr Ile Phe Ile				
	40		45			
ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC						6391
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile						
50	55	60				
TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT						6439
Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile						
65	70	75	80			
ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTT CAATCATGTT AATATAATCA						6496
Ile Ser Phe Lys						
ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAA	ACT					6556
CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAA ACAAGAAAGCA GAGAAC	CATT					6616
AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTACCTA TTGTGATAAT GATGG	TTTT					6676
CTGAGCCCTGT CACAGGGAA GAGGAGATAC AACACTGTT TTATGACCTG CATCT	CCTGA					6736
ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATT	TAAGAA					6796
TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAGAAT GAAGCTAATT ATC	CTTCTAT					6856
ATTCTACAC CTTTGTAAAT TATGATAATA TTTAATCCC TAGTTGTTTT GTT	GCTGATC					6916
CTTAGCTAA GTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTT TTTT	TAATGT					6976
TAATCTAATT GAATAAAAGT TATGAGATCA GCTGAAAAG TAATGCTATA ATT	ATCTCA					7036
AGCCAGGTAT AAAGTATTTG TGGCCTCTAC TTTTCTCTA TTATTCTCCA TT	ATTATTCT					7096
CTATTATTT TCTCTATTT CTCATTATT GTTAGATAAA CCACAATTAA CTATAG	CTAC					7156
AGACTGAGCC AGTAAGAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAG	AGGAGAA					7216
TTCCATGTCA TGAAGACTCT TTTGAGTGG AGATTTGCCA ATAATATCC GCTT	CATGC					7276
CCACCCAGTC CCCACTGAAA GACAGTAGG ATATGACCTT AGTGAAGGTA CCA	AGGGCA					7336
ACTTGGTAGG GAGAAAAAA CCACTCTAAA ATAATATCCA AGTAAGAACA GTG	CATATGC					7396
AACAGATACA GCCCCCAGAC AAATCCCTCA GCTATCTCCC TCCAACCAGA GT	GCCACCCCC					7456
TTCAGGTGAC AATTGGAGT CCCCATTCTA GACCTGACAG GCAGCTTAGT TAT	CAAAATA					7516
GCATAAGAGG CCTGGGATGG AAGGGTAGGG TGGAAAGGGT TAAGCATGCT GT	TACTGAAC					7576
AACATAATTA GAAGGGAGG AGATGCCAA GCTCAAGCTA TGTGGATAG AGG	AAAACCTC					7636
AGCTGCAGAG GCAGATTCAG AAACCTGGAT AAGTCCGAAC CTACAGGTGG ATT	CTTGTG					7696
AGGGAGACTG GTGAAAATGT TAAGAAGATG GAAATAATGC TTGGCACTTA GT	AGGA					7756
GGCAAATCCA TATTGGGGG AGCCTGAAGT TTATTCATT TGATGGCCC TTT	AAATAAA					7816
AAAGAATGTG GCTGGCGTG GTGGCTCACCA CCTGTAATCC CAGCACTTG GG	AGGGCCGAG					7876
GGGGGCGGAT CACCTGAAGT CAGGAGTTCA AGACCAAGCT GACCAACATG GAG	AAACCCC					7936
ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCGTGA AT	CCCACTA					7996
CTCGGGAGGC TGAGGCAGGA GAATCTTTG AACCCGGAG CGAGGAGTT CGAT	GACCT					8056
AGATCGTGCCTT ATTCGACTCC AGCCTGGCA ACAAGAGCAA AACTCGGTCT	CAAAAAAAA					8116
AAAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTTAATACTG TTT	TTAAGTA					8176
GGGCGGGGG TGCTGGAAAG AGATCTGTGT AAATGAGGGA ATCTGACATT TA	AGCTTCAT					8236
CAGCATCATA GCAAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTT GAG	GGGGGGGA					8296
GAGAGTGAGG GGTGGACTAG GACCAGTTT AGCCCTGTC TTTAATCCCT TTT	CTGCCA					8356
CTAATAAGGA TCTTAGCAGT GTTATAAAA GTGGCCTAGG TTCTAGATAA TA	AGATACAA					8416
CAGGCCAGGC ACAGTGGCTC ATGCCTATAA TCCCAGCACT TTGGGAGGGC A	AGGCGAGTG					8476
TCTCACTTGA GATCAGGAGT TCAAGACCAG CCTGGCCAGC ATGGCGATAC T	CCTGTCTCTA					8536
CTAAAAAAA TACAAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAG	CCTACTC					8596
GTGAGCCTGA GGCAGAAGAA TCGCTGAAA CCAGGAGGTG TAGGCTGCAG TG	AGCTGAGA					8656
TCGCACCAACT GCACTCCAGC CTGGCGACCA GAATGAGACT TTGTCCTAAA	AAAAGAAAAA					8716

GATAACAACAG	GCTACCCTTA	TGTGCTCACC	TTTCACTGTT	GATTACTAGC	TATAAAAGTCC	8776
TATAAAAGTTC	TTTGGTCAAG	AACCTTGACA	ACACTAAGAG	GGATTGCTT	TGAGAGGTTA	8836
CTGTCAGAGT	CTGTTTCATA	TATATACATA	TACATGTATA	TATGTATCTA	TATCCAGGCT	8896
TGGCCAGGGT	TCCCTCAGAC	TTTCCAGTGC	ACTTGGGAGA	TGTTAGGTCA	ATATCAACTT	8956
TCCCTGGATT	CAGATTCAAC	CCCTTCTGAT	GTAAAAAAAAA	AAAAAAAAAA	GAAAGAAATC	9016
CCTTCCCCCT	TGGAGCACTC	AAGTTTCAACC	AGGTGGGGCT	TTCCAAGTTG	GGGGTTCTCC	9076
AAGGTCAATTG	GGATTGCTTT	CACATCCATT	TGCTATGTAC	CTTCCCTATG	ATGGCTGGGA	9136
GTGGTCAACA	TCAAAAATAG	GAAAGCTACT	GCCCAAGGAT	GTCCTTACCT	CTATTCTGAA	9196
ATGTGCAATA	AGTGTGATTA	AAGAGATTGC	CTGTTCTACC	TATCCACACT	CTCGCTTCA	9256
ACTGTAACCT	TCTTTTTTC	TTTTTTCTT	TTTTTCTTT	TTTTGAAAC	GGAGTCTCGC	9316
TCTGTCGCC	AGGCTAGAGT	GCAGTGGCAC	GATCTCAGCT	CACTGCAAGC	TCTGCCTCCC	9376
GGGTTCACGC	CATTCTCCTG	CCTCACCCCTC	CCAAGCAGCT	GGGACTACAG	GCGCCTGCCA	9436
CCATGCCAG	CTAATTTTT	GTATTTTAG	TAGAGACGGG	GTTTACCGT	GTTAGCCAGG	9496
ATGGTCTCGA	TCTCCTGAAC	TTGTGATCCG	CCCGCCTCAG	CCTCCCAAAG	TGCTGGGATT	9556
ACAGGCGTGA	GCCATCGCAC	CCGGCTCAAC	TGTAACCTTC	TATACTGGTT	CATCTTCCCC	9616
TGTAATGTTA	CTAGAGCTTT	TGAAGTTTG	GCTATGGATT	ATTTCTCATT	TATACATTA	9676
ATTTCAGATT	AGTCCAAT	TGATGCCAC	AGCTTAGGTT	CTCTCCTAA	ATTGTATATT	9736
GTAGACAGCT	GCAGAAAGTGG	GTGCCAATAG	GGGAACTAGT	TTATACTTTC	ATCAACTTAG	9796
GACCCACACT	TGTTGATAAA	GAACAAAGGT	CAAGAGTTAT	GACTACTGAT	TCCACAACGT	9856
ATTGAGAACT	TGGAGATAAC	CCCGTGACCT	CTGCCATCCA	GAGTCTTCA	GGCATCTTG	9916
AAGGATGAAG	AAATGCTATT	TTAATTG	AGGTTCTCT	ATCAGTGCCT	AGGATCATGG	9976
GAATCTGTC	TGCCATGAGG	CCAAAATTAA	GTCCAAAACA	TCTACTGGTT	CCAGGATTA	10036
CATGGAAGAA	CCTTAGGTGG	TGCCCACATG	TTCTGATCCA	TCCTGCAAAA	TAGACATGCT	10096
GCACAAACAG	GAAAAGTGCA	GGCAGCACTA	CCAGTTGGAT	AACCTGCAAG	ATTATAGTTT	10156
CAAGTAATCT	AACCATTCT	CACAAGGCC	TATTCTGTGA	CTGAAACATA	CAAGAACATG	10216
CATTGGCCT	TCTAAGGCAG	GGCCCAGCCA	AGGAGACCAT	ATTCAAGACA	GAAATTCAAG	10276
ACTACTATGG	AACTGGAGTG	CTTGGCAGGG	AAAGACAGAGT	CAAGACTG	CAACTGAGCC	10336
AATAACAGCAG	GCTTACACAG	GAACCCAGGG	CCTAGCCCTA	CAACAAATTAT	TGGGTCTATT	10396
CACTGTAAGT	TTAATTTC	GGCTCCACTG	AAAGAGTAAG	CTAAGATTCC	TGGCACTTTC	10456
TGTCTCTCTC	ACAGTTGGCT	CAGAAATGAG	AACTGGTCAG	GCCAGGCATG	GTGGCTTACA	10516
CCTGGAATCC	CAGCACTTTG	GGAGGCCGAA	GTGGGAGGGT	CACTTGAGGC	CAGGAGTTCA	10576
GGACCAGCTT	AGGCAACAAA	GTGAGATACC	CCCTGACCCC	TTCTCTACAA	AAATAAATT	10636
AAAAAATTAG	CCAAATGTGG	TGGTGTATAC	TTACAGTCCC	AGCTACTCG	GAGGCTGAGG	10696
CAGGGGATT	GCTTGAGCCC	AGGAATTCAA	GGCTGCAGTG	AGCTATGATT	TCACCACTGC	10756
ACTTCTGGCT	GGGCAACAGA	GCGAGACCC	GTCTCAAAGC	AAAAAGAAAA	AGAAACTAGA	10816
ACTAGCCTAA	GTGTTGGG	GGAGGTCA	ATCGTCTT	GCCGTGAATG	GTATTATAG	10876
AGGACAGAAA	TTGACATTAG	CCCAAAAGC	TGTTGGCTT	TGCTGGAACT	CTACTTAATC	10936
TTGAGCAAT	GTGGACACCA	CTCAATGGGA	GAGGAGAGAA	GTAAGCTGTT	TGATGTATAG	10996
GGGAAACATA	GAGGCCCTGGA	ACTGAATATG	CATCCCATGA	CAGGGAGAA	AGGAGATTG	11056
GAGTTAAGAA	GGAGAGGGAGG	TCAGTACTGC	TGTTCAAGAGA	TTTTTTTAT	GTAACTCTTG	11116
AGAAGCAAA	CTACTTTGT	TCTGTTGGT	AATATACTTC	AAAACAAACT	TCATATATTC	11176
AAATTGTTCA	TGTCCTGAAA	TAATTAGGT	ATGTTTTT	CTCTATAG	GAA ATG AAT	11233
				Glu Met Asn		
			85			
CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG						11281
Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu						
90	95	100				
AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA						11329
Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser						
105	110	115				
TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA						11377
Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys						
120	125	130	135			
CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC						11425
Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe						
140	145	150				
ACT GTT CAA AAC GAA GAC TAGCTATTAA AATTTCATGC C						11464
Thr Val Gln Asn Glu Asp						
155						

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28994 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION: 1..15606
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 15607..15685
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 15686..17056
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 17057..17068
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 17069..20451
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 20452..20468
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 20469..20586
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 20587..21920
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 21921..22054
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: intron
- (B) LOCATION: 22055..26827
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 26828..27046
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: 3'UTR
- (B) LOCATION: 27047..28994
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACTTGCCCTTA	AAAGCTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTTCCCTT	GCTCACATCT	60
GTTCAAGAAA	AATCATTTAA	GTTATAAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
GTTTAGAAT	ATAAACATT	TATACATCAC	CATTTAAATC	TTTCTCCAAG	GCTTCATCTT	180
TATAAAATAG	TCCGGAAATT	TCAGAGAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
GTGGACTATC	TGGCACTGGA	GACTAAATAA	AGAAAGCAGG	TACAGTCAAT	AAGATCTTCA	300
GGACATATAC	ATTTTGTAA	TTAAGAAAAA	GCAAATAAAA	CATTTTCAG	AAAAAGGCAA	360
ACATGCTAGA	AAGCATATGA	CTTAGTCATT	TGAGTTTTA	TTATTAAGGA	AATTTACAGG	420
CCCAAGAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTG	GTGTTCAACT	AGACTTTGCT	480
TTTCATTG	TTGTTTTG	GACAAGTTCT	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
CACAATCTTA	GCTCACTGTA	GCCTCCTAGA	TTCAAGTGAT	CCTCTGTCT	CAGACTCCTG	600
AGTAGCTAGG	ACTACAGGAA	CATTCCACCA	TGCCCGAGCTA	ATTTGTTTT	GTGTTGTTT	660
GTTTTCAGAG	ACAATGTATT	GCAGCGTTGC	CCAGGCTGAT	CTGAAACTCT	TAGCCTCAA	720
CGATACTCT	GCCTCAGCCT	CCCAAAGCAC	TAGGATTACA	GACATGAGCC	AATGCGCCCA	780
GCCTTAAATT	AGACTTAAA	TGTGGTTTA	AACTCCTGTT	GAAAAGCGT	CTGGTATCTT	840
GAACCAGTAG	ATGTTTCAT	AGCAATGAAG	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
CTTGTGAAT	TTTGCTAAAT	AATATAATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
AAGATCAACT	GGTGGGGGCA	GTAGTAAAG	ACAGGATACT	GTGCTTTA	AAAGGTCAGT	1020
AACTATAGTA	CCTAGTTATC	TTACTTATCA	CAGCAAAATA	ATTACATAAA	ATCCTATGGA	1080
TCATAAAAGGC	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
AATAGTTCT	TGGAGGCCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCCCTCC	1200
TATGTTGTAG	GACAATCCTA	GCTCTGGCA	TACGAATACA	TTAAATCCCA	CTTATCTATA	1260

GAGCTTTCTT	AAAGGGAAAGA	AATTTGAGTA	GTATGTAAAAA	CAGAATAAAA	GATTAAGGCT	1320
CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTCTA	TTATCAGGCT	1380
TTACAAGGTG	ATTGCCATC	ATATTCCGAA	GGCACCAAGCT	ACAAAGCTT	GAACAATGCC	1440
AGATTTAGGT	ACAAACTCCA	TGCTACAAGC	TCTCTGGAAT	CCTTCCTCTG	TTCCCACCTCC	1500
TACTGCTGAT	GTAAATTCTAG	ACTGTCATTA	TCTGTCACTT	TCCTAAACTC	AATTTCTCCC	1560
TCCTCTAAAT	CATTCTATCA	ACTGCTATT	GGGTAATCTT	TCAAAACTTT	GATTACTGCA	1620
TTCTCTTAAAC	TCAAAAACCTT	TCATTGTTCC	AGAATAAGTT	GAAATTCCAT	GATATGGCCT	1680
TCAAGGTCTT	GTATTATCTG	GTGCAAGCCT	ACTAGTCCC	TCATTTCAA	CTACTCCTCT	1740
CTATGTACTT	AGCCAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAGG	ACTGGCTCAG	1800
TTAAGATTCT	TTTATCTTCG	GCCGGGCGCG	CTGGCTCACG	GCTGTAATCC	CAGCACTTTG	1860
GGAAGCTGAG	GCAGGAAGAT	CACCTGAGGT	CGGGAGTTCG	AGACCAGCCT	GGCCAGCATG	1920
GTGAAACCTT	GTGCTACTA	AAAATCCAAA	CATTAGCCAG	GCGTGGTGGC	AGGCGCCTGT	1980
AATCCCAGCT	ACTTGGGAAG	CTGAGGTGAG	AGAATCGCTT	GAACCCAGGA	GAGGGAGGTT	2040
GCAGTGAGCC	GAGATTGTGC	CATTGCACTC	CAGCCTGGGC	AACAGAGCGA	GACTCCACCT	2100
CAAAAAAAAGA	AAGGATTCTT	CTATCTCAC	AAAATCTTAA	TGTTAAACA	GGTCTTACAG	2160
TTCATCTAAAT	TCAATCTCAT	TTTTTACAAG	TGAGAAAACA	GGGACAGTGA	CGGTGGATCA	2220
AGTGACACCA	GTAAGACTGA	GCTAAATTAG	AACCGAGATC	TCACTCGAGT	CTGAGGTAT	2280
TCCCCTCTGC	CAACCTTAATC	TTAAAGTAGC	TTCAAATTTT	ACTTTTACTT	TTCCATAAAAT	2340
TCGGAAGGG	TTTCCCTCTAG	GAGTCCAAAT	GTTGAAACCT	GGAAGGGTAT	AGTCTCTGTG	2400
TCTTGAGAT	GAGGGGAGCC	CTGTCATAT	TCAAGTTATC	AATTGACTTT	GTTGTTTTG	2460
AGAACACGATG	CTGATTGGG	TAACTTTAAC	ACATCTGTTT	GATTAGTCCT	ATAAAATATG	2520
CATATATAGA	AGACAGAAAG	AGCAACAACA	AATTGAAAG	ATGCTTGT	AGTAAATTCT	2580
GTATCGTACG	TGTCCATTCC	TGCCAGTACC	TTTATAGTAT	GTAAGTTTAC	GTGCTGTAAT	2640
AGTATTAATA	GTATCTAGAA	AATACTACAC	ATGCACAGCA	GTGCTAACTT	TGCCCTGGGA	2700
GTTGGAAAAT	ACTTCAGAGA	AGCCAACAGG	CAGATTTTC	TCTCTCCCT	TCCCCTTCTA	2760
ATTTTCCCTT	TCCCCTTCAC	CCCCTCTCT	TCTCTCCCCA	AGTAACACTG	TGCACCTATG	2820
TCAAACGAAA	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCTG	TTCCCTGTCA	2880
TGGCTCAAGG	CCTCAGCAGA	TCCAGGCTG	GTGGACGGGC	TGGTCTTCGT	C GTGTGCCAA	2940
ACACTGACCA	CTGGCCCTGGC	TCTGCCATCT	TAGGCTTAGT	GACCTGGCTG	TTACTAAGCA	3000
CTGCTCCCTC	TGCCCATGTC	AGCTGTCCTC	TTCTAGTCTT	CTCCTCTTC	TCAACGCGAT	3060
CCTAGCCCT	CAGGCCATT	CACCTCCATT	TTCCCTCACT	TCECGCCGCC	CCTCCGCACT	3120
TCCTCCCTAC	TGTGTTTCC	GCCCCACTAG	AGCCCCTCAG	AGAAAGTTTC	CATCCTCGCA	3180
CCCTTCCTTG	TGTACAGGCC	CGTCACATTC	TCACAGGCC	CCATCCCTCC	AGCCCCACCC	3240
CAAGGCCAAT	GTACTTCGCG	GTATGGGAC	CTTCCTCGTC	AGCGAACGCG	AGGGAGTGAA	3300
GACCCCTGGC	GCAGGGGTGCT	CGGACTTCGG	GGGTGGAGGT	GGGAAGCGCG	CCGCACCTCCC	3360
AGCAGCCCT	GCACGAGTC	CGTGACAGCT	CTCCCACAC	CACCCCCCCC	AACTTCCCCA	3420
CCGCTAGCTC	CCAGAGCCAG	GCCCCACGGA	AAGGCAGCTT	TTTCCCGGTT	TTCTCCCGCT	3480
CTTCCCCCTC	CACTTGGAAAT	ACTCGTAAA	AAAAAATCTC	TCCCTGCCAC	CCTGTGTG	3540
TTTGAACCAAG	GAAAAAAATCT	GAAACTGGTC	AAGAAAGAAC	AAGGAAGACT	TGCCAAAGCA	3600
AGGCCGGTGT	GTGTCCCAGC	AGCTTAAAGT	CTCAGCAAAG	GAACACAAAAA	TAGCACATCC	3660
ACGGCCTCTT	TTCGAGTAAA	ATTTACTTGG	TTTGTGTTGCA	GGAAGGGTTT	AAAAGTGT	3720
TTGCAGATGC	TCTGTTGCA	GGAAGGCTTT	AATCACGTG	TCCCTGGCC	CACAAGCAAG	3780
GCTTTAGAT	CCAGAGCCTC	AGTTACTGCC	CCCTCTTCCT	CTTGGTGC	ACCAAACGTT	3840
CAGAATCACG	CCTCTTTAGA	AAATTCTTAC	CCCGGGTGTG	TCAATAAGTT	AAGTCTAATT	3900
GGCAACAGCT	ATCAAAAGT	GTTGCATAAC	ACACATGGCT	CACATAATTG	TAGCTTGCC	3960
TCATCGGGTG	TTTAATGCG	GAGGCTTGT	CCTGCAATT	CAAAGATATA	CATTCCAAGC	4020
TTACGCCAG	TTAGTGGATG	TGGAAGAAA	AAAAAAGCAA	ATTACCTCAT	AACACAAAGG	4080
TCAATAAACAC	ACATCCATAA	GCTCCAGGT	CAAAATCTT	CATCTTAGAG	AACTATATT	4140
AAACATTACA	TACATTACTA	AGGTTTTTT	TTCCCTTTG	CTTGATTAAA	TGTTAGTTAT	4200
CATTAAGTCT	TGGAATTATT	CTGTGTGT	ATTTTATT	GCTGTTGT	AAGAAGCCGG	4260
TTGTTTAAAT	TAAGTTCTA	GAAAATAAGC	GCTCAATGTG	TTAATCTGA	GTTGCTAATA	4320
TTGTGAAATA	TAGGCCACAT	AATACTAGCC	TAGATAACTA	TGGCGAAGTA	AGGAGTCTCA	4380
AAACACTGTCC	CAGAACATA	GCAATCTGTG	TTGAATT	ACCCCTGTG	GTAAAATGAA	4440
GGGAAAAGGA	ATGAAGTTT	AGTTGCCTT	AATTTTATC	TTTATTGTTT	CAGACTCTC	4500
AGCAGTATAA	AGTTTCATC	AAGTCAAATA	TATTCACTT	AAAGTGACTG	TGCTTATT	4560
TGATACCAGT	TCCCTCCTAA	TTTGGGGGC	CAGGTGAGAT	AAGTTTATG	AAATAAAAAG	4620
ATTAAAAATT	CTTACATT	TAGTGTCTT	CCTGGTAA	ATGTAGAGTT	GTCCACTGTG	4680
TTTATCTCT	CCTCCTTATT	ATCATGGTG	CTGTTATT	TTTAATGTT	TCATTAAACC	4740
CAAGGGTCTG	GGAAATACTC	ATGGAATTCA	TCTCACAGCC	TTCACACTGT	ATGATATT	4800
AACAGGTGGT	TGTCATCTG	ATTCTAAA	TATTTCAAG	AAAAATGATT	CCACCTAATG	4860
CATAATGCT	TTCATCAGAT	TAAGAGAAC	CCATGGACAT	TTTATT	TTTATT	4920
AAATATTAAC	TTCCATTGCA	TAAGCTAAAT	GGGTAGGAAT	AAGTGAGATG	ATATTGTTAT	4980
CTAGAGCTT	AAAATATTCA	AAGGGCTGTC	ATCATTATCT	CATTAATCT	TTGAAAACAA	5040
CTCTATGAAG	TACAAAGGAC	ACTGAGACAT	TTGTTGCTCT	ATATCAAAGA	AAAAGTGT	5100
TGTCCCAAAA	CTTCAAAATG	TGTAATTAC	ACATTCTGCA	TCTTACAGC	TGGAGAAAAT	5160
TCACTGGCAA	TGGAATATT	AAAATTAGAG	CTTGCTTAGT	GTGCTGCTTC	TGATCACTAC	5220
TTGATCCCAC	TTCGTGCTT	CATGTTAATT	GGCCCAATTG	GACTCTACAG	TTGGAAGGTG	5280

AAAACTTACT	ATTTCAACTT	GAGTCACGTA	TGTATTCTTA	TCATATACTT	CTTAAAGGTA	5340
CTATTTTTT	TCTTCTGATA	GTCACCACAC	CAAGCACTTC	CAGCCACCCCT	GCCACAGACT	5400
TCCTTGAA	TCACTGTTGA	AGGACATGAT	GTTTTATGA	CTTCCGAAA	TGAAAACCT	5460
ATCTTGT	TTAAACAAAC	AAACCAACAA	AAAGTAGTGT	TTATGTAAGC	ATTTGTTCC	5520
CTGACTCTAG	GAACCCCTCT	GTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAAACAA	5580
AATGCCACCT	TGCTAATTCC	CTTCCTAGCA	AAFTAATACA	GTTTAGCACA	TGTTCAAGAA	5640
AAAAATGGCT	AAGAAATTTT	GTTTCCACTA	ATTATTTCA	AGACTGTGAT	ATTTACACTC	5700
TGCTCTTCAA	ACGTTACATT	TTATAAGACT	ATTTTTAAC	ATGTTGAACA	TAAGCCCTAA	5760
ATATATGTAT	CCTTAAATTG	TATTTCAAT	ATTTTAGGTC	AGTCTTGCT	ATCATTCCAG	5820
GAATAGAAAG	TTTAAACACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATT	5880
GGTAGCCCTC	TCCCCAAGCT	TACTTTCTGT	TGCAGAAAGT	GTAAAAATTAA	TTACATAAAA	5940
TTCTAATGAT	GGTATCCGTG	TGGCTTGCAT	CTGATACAGC	AGATAAAGAA	GTTTATGAA	6000
AATGGACTCC	TGTTCCACTG	AAAAGTAAT	CTTAATGGCC	TGTATCACT	ATCCTTGAC	6060
ACCATATTGA	GCTGGGAGG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAGT	AAAAGAAAAT	6120
ATTGCAAA	TGTTCTT	TTTAAATGT	TACATTTCAG	AAATATTTA	AGTGTGTA	6180
CATTGTAGGA	ATTACCCAA	TAGGACTGAT	TATTCCGCAT	TGTAAAATAA	GAAAAAGTTT	6240
TGTGCTGAAG	TGTGACCAGG	AAGTCTGAAA	ATGAAGAGAG	ACAGATGACA	AAAGAAGATG	6300
CTTCTAATGG	ACTAAGGAGG	TGCTTCTTA	AAAGTCAGAAA	GAGACTCTCA	GAAAGAGGTA	6360
CAGGTTTTGG	AAGGCACAGA	GCCCCAACTT	TTACGGAAGA	AAAGATTCA	TGAAAATAGT	6420
GATATTACAT	TAAAAGAAGT	ACTCGTATCC	TCTGCCACTT	TATTCGACT	TCCATTGCC	6480
TAGGAAAGAG	CCTGTTGAA	GGCGGGCCA	AGGAGTGC	ACAGCAGTCT	CCTCCCTCCA	6540
CCTTCTTCT	CATTCTCTCC	CCAGCTTGCT	GAGCCCTTTG	CTCCCTG	GACTGCCTGG	6600
ACAGTCAGCA	AGGAATTGTC	TCCCAGTGCA	TTTGCCCTC	CTGGCTGCCA	ACTCTGGCTG	6660
CTAAAGCGGC	TGCCACCTGC	TGCACTAC	ACAGCTTCGG	GAAGAGGAA	GGAACCTCAG	6720
ACCTTCCAGA	TCGCTTCC	TCGCAACAA	CTATTGTCG	CAGGTAGAGA	ATATCATTCC	6780
TCTTATTG	GAAAAGTACG	CATGGCAATT	AGAGGTAAAT	AAGCTAGAAA	GCAATTGAGA	6840
GBAATATAA	CCATCTAGCA	TCACTACGAT	GAGCAGTCAG	TATCACACATA	AGAAATATAA	6900
GCAAAGTCAG	AGTACAATT	TTTCTTTA	TCAGATATGG	GAGAGTATCA	CTTGTAGGAA	6960
GAGGTTCTCA	AACTTTTG	TCTCATGTT	CCTTACACT	AAGCACATCA	CATGTTAGCA	7020
TAAGTAACAT	TTTAATTAA	AAATAACTAT	GTACTTTTT	AAACACAAA	AAAAGCATAA	7080
AGAGTGACAC	TTTTTATT	TTACAAGTGT	TTAACTGGT	TTAATAGAAG	CCATATAGAT	7140
CTGCTGGATT	CTCATCTGCT	TTGCATT	ACTACTGCAA	TATTGCACAG	AATGCAGCCT	7200
CTGGTAAACT	CTGTTGTACA	CTCATGAGAG	AATGGGTGAA	AAAGACAAAT	TACGTCTAG	7260
AATTATTAGA	AATAGCTTC	ACTTTAGGAA	CTCCCTGAGA	ATTGCTGCTT	TAGAGTGGTA	7320
AGATAAATAA	GCTTCTT	AAACGGAATC	TCAAGACAGA	ATCAGTTACA	TTAAAAGCAA	7380
ACAAAAAAAT	TGCCCATGGT	TAGTCATTT	TGAAATCTG	CCACACCTT	GGACTGGGCT	7440
ACAATTGGAT	AATATAGCAT	TCCCCGAGAT	AAATTCTCT	CACAATTAAG	GAAAGGGCTG	7500
AATAAATATC	TCTGTTGAA	GTTGAATAAC	AAAAATTAGG	ACCCCTAAA	TTTTAGGGCT	7560
CCTGAAATT	GTCTTTTG	CTATATT	CTACTTTACG	TTCTATTAAA	TCTTCTTCA	7620
GGCCAGGTG	ACTAGCTCAT	GCCTAGAATC	TCAGGCAGGC	CTGAGCCCAG	GAATTGAGA	7680
CCAGCCAGGG	CAACACAGTC	TCTACAAAAA	AATAAAAAT	TACCTGGGTG	TGTTGGTCA	7740
TGCCTGTAGA	ACTACTCAGG	ATGCTGAGGA	CTGCTTGAGC	CCAGGATAGC	CAAATCTGTG	7800
GTGAGTTCAG	CCACTAAACA	GAGCGAGACT	TTCTCAAAA	AAACACAAA	AAAACAAACA	7860
AACTCCCTC	AAAATAACTT	TTTATCTGCA	ATGTTTCCT	ATTGCTG	AGATTAATT	7920
TACTCTT	CCTGATT	AAAGCCCTC	ATAATCTAAT	CCGACTTTAC	CTTGTGTTCA	7980
CTGCAAAATA	GCAGGACTGT	TCCACTAAC	TCCAAAATC	ACAGGTTGGG	TGCACTGGCT	8040
CACTCCTGTA	ATCCAAACAC	TTTGGAGG	CAAGGCAGGT	GGATTGCTTC	AGCTCAGGAG	8100
TTCAGACCA	GCCTGGCAA	CATGGCAAAA	ACCCGTCTC	TCCAAAACAT	ACAAAAATTAA	8160
GCCAGATGTG	GTAGTATGTG	CCTGTAGTCC	CAACTACTCA	AAAGGCTAAG	GCAAGAGGAT	8220
CACTTGAGCC	CAGGAGGTCA	AGGCTACAGT	GAGCCATGTT	TACTGTGTC	CTGCACTCCA	8280
GCCTGGTGA	TAGAGCAAGA	CCATGCTCA	AAAAAAA	AAAGAAAAGA	AAAGAAAAAA	8340
ACATCGCTCT	ATTCAAGTTCA	CCCCCACCAC	AAACATTGTT	TGATTATCAC	ATAAAATGCTG	8400
GTCCATTGCC	TTCTCTATCT	ATTCAAAATCT	TTAAGCATTC	TTTGAGATT	AACTCAATT	8460
TCCTTTCAA	ACTAGGCCAT	TTAAACTACA	TCAGTCCAT	TTTGATT	TTGCTT	8520
TCTACAGACT	CAAAACAAA	AACTTAAA	CTTATTTT	AAAGTTTCTG	CTACTCTCAC	8580
TTCTCAACAA	CTCACATACA	CGCATT	ATAAGATG	AGAATGTTCA	AGGATAAAAT	8640
GATTATAGA	ACTGAAAAGT	AGGTTTG	TCTTGTG	GTCAAGATGA	CTACCTACCT	8700
GATCTCAGGT	ATTAAATTAT	GTAGCATG	CCCTCAT	ATCCCAT	TATTCAACAG	8760
GATTGGAATT	CCACAGCAAG	GATAAACATA	ATCATAGTG	CTTTCAAGT	TCAAGGCATT	8820
TTAACCTTTA	ATCTAGTAGT	ATGTTGTTG	TTGTTGTTG	TGTTGAGAT	GGAGCCCTGC	8880
TGTGTACCC	AGGCTGGAGT	GCAGTGGC	GAAC	CTGCAACC	TCTGCCTCAT	8940
GGGTTCAATC	AGTATTCTG	CCTCAGTG	CCAAGTAGCT	GGGACTACAA	GGCACATGCC	9000
ACCATGCC	GCTAATT	GTATTTAG	TAGAACAGG	GCTTCACC	GTTGCCAGG	9060
CTGGTCTCGA	ACTCCTGACC	TCAAGT	CAGCCGC	GGCCTCC	AGTGTGG	9120
TTACAGGCA	AAGCCACCGT	GCCCAGC	ATAGTATG	TTAAACTCT	TAGTGGCTTA	9180
ACAATGCTGG	TTGTATAATA	AATATGCCAT	AAATATT	TGTCTTAGAA	TTATGAAGAA	9240
GTGGTTACTA	GGCCGTTG	CACAT	TGGTCTC	CTTACAGCTT	TAATTAGAGT	9300

CTAGAATTGC	AGGTGGTAG	AGCTGGAACA	GACCTTAAAG	ATTGACTAGC	CAACTCCCT	9360
GTCCAAATGA	GGGAACTGAG	ACCCTAAAAA	TTAAGTGACT	TGCCCGAGAC	AAAACTGGAA	9420
CTCATGTGTC	CTAACTTCCA	TCATGAAATT	CTACCATCA	CTAGCCTCTG	GCTAGTTGTC	9480
AAAGTATTGC	ATAACTAAAT	TTTTATGTC	GTGTTAAAGA	ACAAATTGTC	ACTGCTTACT	9540
CCTGGGAGGG	TCTTCTGAG	GTGGTTATA	ACTCTTAAAA	AAAAAAAAGT	CAGTAGTCTG	9600
AGAATTTTAG	ACGAAATAGT	CAAAGCATT	TTATCCAATG	GATCTATAAT	TTTCATAGAT	9660
TAGAGTTAAA	TCAAAGAAC	ACGGATGAGA	AAGGAAGAGG	AAAATTGAGG	AGAGGAGGAA	9720
TGGGGATGAG	AACACACTAC	TTGTAATCAG	TCATAGATGT	ACTGAGAACT	AACAAGAAGA	9780
ATTGTAAGAA	AATAAGAATG	AAGAATTCAA	AATCAACACA	TGAAATAAAA	AGAAACTACT	9840
AGGGAAAAT	GGAGAAGACA	TTAGAAAAAT	TATTCTATT	TTAAAATTCT	GTTTCAGGC	9900
TTCCCCTCTG	TTCTCCTCC	TTCTCATTGG	TTTCAGGTG	GAGGGAAAGT	TTAAGATGGA	9960
AAAAATATAT	ATATTCTACA	CATCCCTTC	TACGCTGTTG	TCATGGCAC	AAGGTTTATC	10020
ATAGCAAAC	TTTATTCTATA	CAACATTAT	TGAGTCTTA	CTGTTGGTA	AGCTCTTCC	10080
AGGTGTTGAA	AATTCAAGGG	AAAAAGACAA	ACTCATTGTC	TTAAAACCTA	GATGAAAGCT	10140
GAACAGACCT	ATTTTAATC	AAAGTAATCT	CAATTAGGG	TAGTAAGAGC	TATTTAAGAA	10200
GCATGAACAG	GTGTGAAGGA	GGTAGGACTC	TGAGGAGAGA	ATAGTTAGCT	AGGAATGAAA	10260
GAGCAGAGAA	GTTCCTCTAG	AGGAACATT	AAAGCTGGGA	GTTACGGGAT	GAAAGATGAG	10320
GCAGGGTTG	CAGGAAAAAA	AAAAAAAAG	GCAGGGGAAG	GGGAAGTTCT	GGCCTGGCAG	10380
AGAGAATAAC	TGTGGCAACA	ATGGAGGAGA	GTCTGGAAGC	AAGAAAACCA	AGTAGAAGAG	10440
TATTAATAA	GAAGATGCCA	GGGGTAATGA	GGGCTTGATT	AAAACAGTG	CTGTTGGAGA	10500
TGGAGAGGAG	ATACCAAATT	CTGGAGACAT	TTCTGAGTTA	GAACCTACAG	TATTTATCAG	10560
ACAAGGGAAA	GATTAGACAA	AGGAGTTAAG	AATGACTCCC	AGGTTTCAGT	TTGGGGCAGG	10620
TAACTAGGAC	ATGTTTTGAA	AAGTAATGTA	TTGGATCTCT	TACCATGGA	ACTATGTATG	10680
TGGAGCCAAA	TTAAAATTG	TACATGTATA	TAACTCTCCC	CCCACACCCA	GTAACTACTT	10740
CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAAGAAT	AGTTTGTAGT	CACTGTCTT	10800
ACTTTCCCC	TCCCATTCTG	TCCTAGATAT	TTGTCACCT	ACCATCTGCT	GCCTCCACTT	10860
TACCCAAAC	GTTCTACGGT	TGCCCCAAAC	TTCTTAATTG	CCAAATTCAA	TGAACAAGTT	10920
TAAGCTTATA	TGTAATTAG	GAGCTCTACA	GTTCGATTTC	GAGCAGCCCC	TCCTGAAACC	10980
CTTTCTCTT	CGACTTCTGT	GACACATCTC	AGATTACAA	AACTGAACTA	ATTATTTAC	11040
ACTTGAGCTG	TATTTCGTT	CTTCTTCTT	GATGAATGAG	GTAACCACTC	AACAAATTGC	11100
CCAAGCCAAA	AACTACGAAG	TCATCCTCAG	TTCTCTCTTC	TTCTGTTTG	CCCACAAACAG	11160
ATCAGCTGAG	AAATCCCGCT	GTTTAGTATC	TCTTGAATT	ATTACCTTAA	TTTATAGCCT	11220
CATCAACTCT	TAATTGTTAA	AATTACCTCA	GTAGTTGTG	TCTGACCTCT	GTCCAATCTT	11280
GTTCATCTAG	GTCCATTCTT	TTGTTCTTGG	TGGTGGTGGT	GGTGTTGACA	GAGTTTCGCT	11340
TTTGCTGCC	AGGCTGAAGT	GCAGTGGAGC	ACTTCACTGC	AACCACAGCC	TCCTGGGTTT	11400
AAGCAGTTCA	CCCTCCCGAG	TAGCTGGAC	TACAGGTATG	TGCCACACCA	CCCAGCTAAT	11460
TTTGTGTTT	CAGTAGAGAC	AGGGTTTAC	CATGTTGGTC	AGGCTGGTCT	CAAACCTCTG	11520
ACCTCAAGCA	ATCCACCCAC	CTCAGCCTCC	CAAAGTGTG	GGATTACAGG	CATGAGCCAC	11580
TGCACACCGA	CCAGATCCAT	TGTTTATGTT	GCTCTAGAG	TGAGTTTTA	AAACACAAAT	11640
TTGACCATAT	CTTCTCCAA	TTAAGTCAG	TATTTTTTT	TTCAAGGAAAA	AAACAGTTCAA	11700
ACTCTTTAGT	CTGCTTACAC	AAGGCCCTTG	TAGTCTGACT	CTTCTTCCA	AGCTTTCATC	11760
AAAGTATACT	GCAAGTTACA	TTTTATGTA	ATTGAATTAG	GCAACGGTAT	AAAATTATA	11820
GTTTATATGG	GAAAATGGA	AATAATGTTA	ACTCTTCAA	ATAGTTTATC	TAGAATGACA	11880
TAATTTCAAA	GCTGTCAGGT	CAAATGAGTT	ATAAACTGTT	AAACACTATTG	CCACATGCAA	11940
GTGTCTCTTA	TACTTGGTAG	AATTATCTG	TCCTCATGTCA	TTATTATGTA	AATTAGACTT	12000
TAAATAACTC	AGAAGTCTT	CAGACATACA	GGTTTATT	GTGCTTTTA	AAACATAATT	12060
TAAATAATT	TATATATGAT	AATGTTATCC	AAAGTGTAG	GGATGTTATTG	TTACTGCTGT	12120
GCAAAAAAAA	AAAAAAAAT	AACTCCAAAT	AAATATGTG	AAACCAAGTT	TATATGCAAG	12180
AAAACAATAT	AAAAAAGGCC	AAAGTACAC	CATAATAGGC	TGTGTGGAGA	CGGCAGGCTA	12240
CAAAACACTA	GTAATAATGC	TGAGAAAGTT	AAAAAAAGAA	AGAAAGCAAC	AATATGCTT	12300
GGTTGTTGTA	GGTTTATGTA	CTCCAAGAAT	ATCTCCTCTC	AAACTTTAC	GTTTTTCCA	12360
AAGAAAAGTT	AACTTTGGCT	GGGCGAGTG	GCTCTTGCT	GTAGTCCCAG	CCTTGGGAG	12420
GCCAAGGGCG	GCAGATCACC	TGAGGTCAAG	AGTTTGAGAC	CAGCCTGACC	AAAATGGAG	12480
AAACCCGCC	CCCTCACTAC	TAAAAGAATA	CAAATTAGG	CGGGCAGACAG	TGGCTTACCC	12540
CTGTGATCCC	AGCACTTGG	GAGGCCGAAG	CAGGAAGATC	ACCTGAGGTC	AGGAGTTCGA	12600
GACCAGCCAT	GGAGAAACCC	GTCTCTACTA	AAAATACAAA	ATTAGCCGGG	CCTGGTGGTG	12660
CATGACTGTA	ATCCCAGCTA	CTCAGGAGGC	TAAGGCAGAG	AATCACTTG	ACCCAGGCAG	12720
TGGAGGTTGC	AGTGAGCGA	GATCGTGC	TTGCACTCCA	GCCTGGGCAA	CAAGAGCGAA	12780
ACTCTGTATC	CAAAAACAA	AAGAAAAGAA	AAGGTAACCT	TGAACATATGT	GAGATCTTA	12840
GAAATGCAATT	CTTCTGTAA	AATGTGACTA	CATTGCTT	ATTATGGTA	AAAATGTTGA	12900
GGCCTAAAC	AACCCATATT	TTCTCGGTCT	CCCCGCTGCC	TAGCCTTTGT	TCACATTGCT	12960
TCTTCTGGT	GGAAAGCTCTT	CCTCTGGCCT	TGAAATGCC	TGCTTCTCTT	TCAAGGTTAGC	13020
ACAGTCATCA	CTTTCTGTGG	TAACCTTCTC	CAGCACCAC	AAACAGAAAG	AATGAATCTC	13080
TTGTAAATT	AGCTCTTACG	TCATTCTTCA	CATTATTTG	TAACCTTTA	TAGATTCTTC	13140
TCTCCCACTA	GACTCTGAGT	CACTGGAGAG	TAGGAGCCA	CTCTCATTCA	TGTGTGGTTT	13200
GGTCAGCTAC	TGGCCACATT	CCTGATGCAT	AGTTAATGCT	CAAACCTTAA	CTGGTGAATC	13260
AGCTCAAATA	TTGTCCTTCT	CTAAATCCAT	TCACTCATTG	ACTAACTATG	TACTCAAAT	13320

AGTAAACACC	AGTAATTAA	TCCAATTCT	GCCCATACTG	CTGGTACAT	TTCAGGTGAA	13380
TTAGTTGAT	AAATATGTT	GTATTACATA	ATATTAAAGT	ATGTACAGAA	GATCATGCTA	13440
ATCATAATTC	ACAATGATA	ACTAACAAA	CATAAATGC	CTCAGGTTAA	CAAATGCTG	13500
CCTTCTCACT	TAATGCAGTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGGCCTTGT	13560
CAGCAATGAA	GCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
ATTATTGTTG	TGATTTTAA	TATTAATGTT	GGTTGGGTT	TTGGTGAAA	ATGTTTAGAC	13680
TTATCTTAGT	GATCTTTCA	TCCTTGCTA	TATTATTTT	CTCTAAGAGT	CTTCCTTATC	13740
CCCTCCTTAA	AAAAACTAGG	TGATAATTCT	AAATTGTA	TTAAATATT	ATAAAATAGCT	13800
TATAAAATT	AATATTATA	ATATTTAAAT	GTGATAAA	TATTTAAATT	TTATAATATT	13860
TAAATGTTA	TTAAATTCA	TTGTACATC	AGTTTTATT	TTATTAAAT	GTGTTGGCCA	13920
GGCATGGTGG	CTGACACCTA	TAATCCAGA	ACTTTGAGAG	GCCAAGTCAG	GCAAACCAT	13980
TGAGCTCAGG	AGTTGAGAC	CACCCCTGGC	AACGTGGTGA	AACCCGTCT	CTACCAAACA	14040
TATGAAAATC	TATCTGGGTG	TGGTGGCACG	CATCTGTGGT	CCCAGATGGG	AGTCCCAGGC	14100
TAAGATGGGA	GAATCGCTG	AACCCAGGTG	AGAGGGTGG	GGTGGATGTT	GCAGTGAGCT	14160
GAGATCGTC	CACTGCAC	CAACCTGGGT	GACAGAGTGA	GACTCCATCT	CAAAAAAAA	14220
AAATGTTATC	AAATAAAGAT	AAATTTATA	ACTGTTCGCA	CTTAGATGAG	CATAAGGAAC	14280
TAAACCTAGA	AAAACATATC	AAATAAGGCC	TGGGTACAGT	GACTCATGCC	TGTAATCTCA	14340
AGCACTTGG	GAGGCCAAA	TTATACAAAG	TTAGTTGTAT	AACACCAACT	AACAACATT	14400
TTGGGGTTAG	CTTAATTCA	ATTAATT	TTAAACTGA	GTGTTAAATT	CCTGCTTACT	14460
CTACCATACA	TGCTAGGCCT	CATATTATGC	TAGAAAAATT	TTGAGCACAG	ATTTATGAAT	14520
ACTCTCCTGC	ATACCATTAA	ATTTTAAAC	AAATTAAAT	GCAGTATATA	TGTGCCTTT	14580
TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGTA	TTTCAAAGTA	14640
GTAATGAGTT	TAACCATGT	CTCAAGATCT	CTGCAATAAC	TGAGCACAA	CAGAAAATAG	14700
GTATTCTAT	TAATGACAGA	GTCACAAGTA	CTACTAATAA	TACTGTGGTT	TGTTCCCTGC	14760
AAACTATCAT	GGGAGGAATG	CTAAATTCA	GAGGTTGGT	AAAATACATG	TGTATT	14820
TCCCCATCCA	AGTCAACAGA	TTTCTCACAC	TGAGAACCTC	TATTCATAA	CAAAATTCTG	14880
GAAGCCTGCA	CACCGTATTG	GAAGAAGGGC	AGAAGGGAA	AGCAATGGA	AGGATTTAA	14940
TTTTTTCAA	ATCCTGTATC	CCTTGATT	ACAGCAAGAT	TGTATTATG	TATTACTGT	15000
GTTAAAATA	TAGTATAATC	GAGACTCCAG	ATCAAAATC	ACCGCAGCTC	AGGGAGAAAG	15060
AGGGCCACCA	AATGCCAGAG	CCCTTCAGCC	TTCTCCCACC	CTGCCTGTAC	CCTCAGATGG	15120
AAGCACTTT	TTATCATTGT	TTCACCTTA	GCATTTGAC	AATGAAGTC	CAAACCTTCA	15180
GCCTCTCACC	CATAGGAACC	CACTGGTTGT	AAGAGAAGGA	TGAAGCCAGT	CCTTCCTAAA	15240
GGGCACGATT	AGATGTGTTT	ATGGCATTCT	CAGGTGAAAC	TATATTATA	TTGACAATAT	15300
ATTTATATT	CTCAAGGAAT	ACTAGAATAA	TGATTCA	CGACTACTAGG	CCATTATCT	15360
ACCCCTTATA	ATATTGTTA	ATGAGAAAAT	GCTTCTATC	TTCCAAATAT	CTGATGATT	15420
GTAAGAGAAC	ACTAAACAT	GGGTATTCTAT	AAGCTGAAAC	TTCTGGCATT	TATTGAATGT	15480
CAAGATTGTT	CATCACTATC	CTAGGTGATT	AACTGACCAC	TGAACATTGAA	GGTAGTATAA	15540
AGTAGTAGTA	AAAGGTACAA	TCATTGTCTC	TTAACAGATG	GCTCTTGCT	TTCATTAGGA	15600

ATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA 15651  
 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala

-35 -30 -25

ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGGC TAATGCCATA 15702  
 Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala

-20 -15 -10

GAACAAATAC	CAGGTCAGA	TAATATCTATT	CAATTAGAAA	AGATGTTGTG	AGGTGAACTA	15762
TTAAGTCACT	CTTGTGTCA	CCAAATTTC	CTGTAATATT	AATGGCTCTT	AAAAAAATAG	15822
TGGACCTCTA	GAATTAACC	ACAACATGTC	CAAGGTCTCA	GCACCTTGT	ACACCACTGT	15882
TCCTGGCACT	TTATCAGCA	GTAGCTCACT	CTCCAGTGG	CAGTAAGTGC	ACATCATGAA	15942
AATCCCAGTT	TTCATGGGAA	AATCCCAGTT	TTCATTGGAT	TTCCATGGGA	AAAATCCCAG	16002
TACAAAATCG	GGTGCATTCA	GGAAATACAA	TTTCCCAAAG	CAAATTGGCA	AATTATGTAA	16062
GAGATTCTCT	AAATTTAGAG	TTCCGTGAAT	TACACCATT	TATGTAATAA	TGTTTGACAA	16122
GTAAAAATTG	ATTCTTTTT	TTTTTTCTG	TTGCCAGGC	TGGAGTGCAG	TGGCACAATC	16182
TCTGCTCACT	GCAACCTCCA	CCTCCGGGT	TCAAGCAATT	CTCCCTGCCTC	AGCCTCTGA	16242
GTAGCTGGGA	CTACAGGTGC	ATCCCCCCAT	GCCTGGCTAA	TTTTGGGTA	TTTTTACTAG	16302
AGACAGGGTT	TTGGCATGTT	GTCCAGGCTG	GTCTTGGACT	CCTGATCTCA	GATGATCCTC	16362
CTGGCTCGGG	CTCCCAAAGT	GCTGGGATTA	CAGGCATGAA	CCACACACAA	TGGCCTAAAA	16422
ATTGATTCTT	ATGATTAATC	TCCTGTGAAC	AATTGGCTT	CATTGAAAG	TTTGCCTTCA	16482
TTTGAAACCT	TCATTTAAA	GCCTGAGCAA	CAAAGTGGAGA	CCCCATCTCT	ACAAAAAAACT	16542
GCAAAATATC	CTGTGGACAC	CTCCTACCTT	CTGTGGAGGC	TGAAGCAGGA	GGATCACTTG	16602
AGCCTAGGAA	TTTGAGCCTG	CAGTGAGCTA	TGATCCCACC	CCTACACTCC	AGCCTGCATG	16662
ACAGTAGACC	CTGACACACCA	CACACAAAAA	AAAACCTTCA	AAAAAAATTA	TTAGTTGACT	16722
TTTCTTAGGT	GACTTCCGT	TTAACGAAATA	AATTAAAAG	AAAATCTCT	AATTTTAGAA	16782
AATTATTTT	TAGTTACATA	TTGAAATT	TAAACCTCTAG	GTTTAAGTTT	TATGTCTAAA	16842
TTACCTGAGA	ACACACTAAG	TCTGATAAGC	TTCATTTAT	GGGCCTTTG	GATGATTATA	16902
TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCATA	AAAATAGGAG	TTCGAGAAAG	16962
AGGAGTAGCA	AAAGTAAAAG	CTAGAATGAG	ATTGAATTCT	GAGTCGAAAT	ACAAAATTTT	17022

ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAGT 17075  
 Ala Glu Asp Asp Glu  
 -10

AGAAATGAAT TTATTTTCT TTGCAAACATA AGTATCTGCT TGAGACACAT CTATCTCACC 17135  
 ATTGTCACT GAGGAAAAAA AAAAATGGTT CTCATGCTAC CAATCTGCCT TCAAAGAAAT 17195  
 GTGGACTCAG TAGCACAGCT TTGGAATGAA GATGATCATA AGAGATACAA AGAAGAACCT 17255  
 CTAGAAAAG ATGCTTCTCT ATGCCCTAAA AAATTCTCCA GCTCTTAGAA TCTACAAAAT 17315  
 AGACTTTGCC TGTTTCATTG GTCCTAAGAT TAGCATGAAG CCATGGATTG TGTTGTAGGG 17375  
 GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAATC AGTAACACCT 17435  
 CCTCTCAGAA ATGCTTGGG AAGAACGCTG GAAGGTTCCG GGTTGGTGGT GGGGGGGGC 17495  
 AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGC AAGAACGACCA CATTCAAGAGG 17555  
 CCAAAAGCTG AAAGAAACCA TGGCATTAT GATGAATTCA GGGTAATTCA GAATGGAAGT 17615  
 AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGTAAACAGGGTG TAGAGCAAGA 17675  
 CGTTCTCTCA CCCCAAGATG TGAAATTGG ACTTTATCTT GGAGATAATA GGGTTAATTAA 17735  
 AGCACAATAT GTATTAGCTA GGGTAAGAT TAGTTTGTG TAACAAAGAC ATCCAAAGAT 17795  
 ACAGTAGCTG AATAAGATAG AGAATTTC TCTCAAGAA AGTCTAAGTA GGCAGCTCAG 17855  
 AAGTAGTATG GCTGGAAGCA ACCTGTGAT ATTGGGACCC CCAACCTCT TCAGTCTGT 17915  
 ACCCATCATC CCCTAGTTGT TGATCTCACT CACATAGTTG AAAATCATCA TACTCCCTGG 17975  
 GTTCATATCC CAGTTATCAA GAAAGGGTCA AGAGAAGTCA GGCTCATTCC TTTCAAAAGAC 18035  
 TCTAATTGGA AGTAAACAC ATCAATCCCC CTCATATTCC ATTGACTAGA ATTTAATCAC 18095  
 ATGGCCACAC CAAGTGAAG GAAATCTGGA AAATATAATC TTTATTCCAG GTAGCCATAT 18155  
 GACTCTTAA AATTCAAGAAA TAATATATT TTAAAATATC ATTCTGGCTT TGGTATAAAG 18215  
 AATTGATGGT GTGGGGTGAG GAGGCCAAA TTAAGGGTTG AGAGCCTATT ATTTTAGTTA 18275  
 TTACAAGAAA TGATGGTGTC ATGAATTAAG GTAGACATAG GGGAGTGTG ATGAGGAGCT 18335  
 GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTAGG GTTGGATTG 18395  
 GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTT CACTTAGGCA ATTGTACCA 19455  
 TTCAGTAAA TAGGAAACAC AGGAGGAAGA GCAGGGTTG GTGTATACAA AGAGGAGGAT 18515  
 GGATGACGCA TTTCGTTTG GATCTGAGAT GTCTGTGGAA CGTCTAGTG GAGATGTCCA 18575  
 CAAACTCTTC TACATGTGGT TCTGAGTCA GGACACAGAT TTGGCTGGA GATAGAGATA 18635  
 TTGTAGGCTT ATACATGAA ATGGCATTG AATCTATAGA GATAAAAAGA CACATCAGAG 18695  
 GAAATGTGTA AAGTGAGAGA GGAAAGCCA AGTACTGTGC TGGGGGAAT ACCTACATT 18755  
 AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC AGACTAACCA AAAGGGGAGA 18815  
 AGAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG 18875  
 TGTGTGTTG AATTTCAG CTTGAGAAT CAAGGGCCAG AACACAGCTT TTGAGTTAG 18935  
 CAACAAGGAG TTTGGTGTAC TCAGTGAAG CAGCTGTGATG GTGAAATGGA GGCAGAGGCA 18995  
 GATTGCAATG AGTGAACACAG TGAATGGAA GTGAAGAAAT GATACAGATA ATTCTTGCTA 19055  
 AAAGCTTGGC TGTAAAGG AGGAGAGAAA CAAGACTAGC TGCAAAAGTGA GATTGGTTG 19115  
 ATGGAGCAGT TTTAAATCTC AAAATAAGA GCTTGTGCT TTTTGATTA TGAAAATAAT 19175  
 GTGTAAATG TAACTAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAATAAT 19235  
 AAAAACCAAC CAGAAATAAT GATAGCTACC ATTGTGATAC AATATTCTA CACTCCCTTC 19295  
 TATGTATATA TACAGACACA GAAATGCTTA TATTTTATT AAAAGGGATT GTACTATACC 19355  
 TAAGCTGCTT TTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG 19415  
 CAGTTATATT AAGTCATGA TATTCACAA TAAGGGCATA TCTTGCCCT TTTTATTAA 19475  
 TCAATTCTTA ATTGGTGAAT GTTGTGTTCC AGTTTGTGTT TGTTATTAAC AATGTTCCCA 19535  
 TAAGCATCTC TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTCTTCA GGATAAAAACC 19595  
 CAGGAGGTAG AATGCTGGG TTGATAGAAG AGAAAGGGATG ATTGCCAAAT TAAAGCTCAG 19655  
 GTAGAGGAGTA CATGCCGAGC ACAAATGGGA TCAGCCCTAG ATACAGAAA TGGCAGTTTC 19715  
 TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTG 19775  
 TACGTGGAGT AGCAGGAAT CATTGCTGA AAATGAAAAG AGAGATGATG TTGTAGAGGT 19835  
 CCTGAAGAGA GCAAAAGAAA TTTGAATTG CGGCTATCAG CTATGGAAGA GAGTGTGAA 19895  
 CTGGAAAACA AAAAGAGTAT TGACAAATTG TATGCTTGTG ATGGCACCGA TTTGAACGCT 19955  
 TGTGCCATTG TTCAACAGCA GCACTCAGCA GCCAAGTTG GAGTTTGTGTA GCAGAAAGAC 20015  
 AAATAAGTTA GGGATTAAAT ATCCTGGCCA AATGGTAGAC AAAATGAACCT CTGAGATCCA 20075  
 GCTGCACAGG GAAGGAAGGG AAGACGGGA GAGGTTAGAT AGGAAATACA AGAGTCAGGA 20135  
 GACTGGAAGA TGTGTGATA TTTAAGAACATAGAGTTG GAGTAAAAGT GTAAGAAAAC 20195  
 TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTGAA GTTAACACTT CAGAGGCAGA 20255  
 GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCCTTGA GAGTAGGTTA AAAAACAAATA 20315  
 GGCAACTTTA TTGTAGCTAC TTCTGGAAACA GAAGATTGTC ATTAATAGTT TTGAGAAACT 20375  
 AAAATATATA GCATACTTAT TTGTCAATTAA ACAAAAGAAC TATGTATTT TAAATGAGAT 20435  
 TTAATGTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA 20486  
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu

-5                    1                    5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534  
 Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile  
 10                 15                 20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT 20582  
 Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

	25	30	35	
AGA G	GT ATTTTTTTA	ATTCGAAAC	ATAGAAATGA	CTAGCTACTT CTTCCCATTC 20638
Arg Asp				
40				
TGTTTACTG	CTTACATTGT	TCCGTCTAG	TCCCAATCCT	CAGATGAAAA GTCACAGGAG 20698
TGACAATAAT	TTCACTTACA	GGAAAATTTA	TAAGGCATCC	ACGTTTTTA GTTGGGGTAA 20758
AAAATTGGAT	ACAATAAGAC	ATTGCTAGGG	GTCATGCCCT	TCTGAGCCTG CTTTGAAATC 20818
ACCAATCCCT	TTATTGTGAT	TGCATTAAC	GTTAAAACC	TCTATAGTTG GATGCTTAAT 20878
CCCTGCTTGT	TACAGCTGAA	AATGCTGATA	GTTTACCAAGG	TGTGGTGGCA TCTATCTGTA 20938
ATCCTAGCTA	CTTGGGAGGC	TCAAGCAGGA	GGATTGCTTG	AGGCCAGGAC TTTGAGGCTG 20998
TAGTACACTG	TGATCGTACC	TGTGAATAGC	CACTGCACTC	CAGCCTGGGT GATATACAGA 21058
CCTTGTCTCT	AAAATTAAAAA	AAAAAAAAAA	AAAAAACCTT	AGGAAGGAA ATTGATCAAG 21118
TCTACTGTGC	CTTCACAAAC	ATGAATTCCA	AATATCAAAG	TTAGGCTGAG TTGAAGCAGT 21178
GAATGTGCAT	TCTTAAAAAA	TACTGAATAC	TTACCTAAC	ATATATTTTA AATATTTTAT 21238
TTAGCATTAA	AAAGTTAAAAA	ACAATCTTT	AGAATTCTATA	TCTTAAAAT ACTCAAAAAA 21298
GTTCAGCGT	GTGTGTTGTA	ATACACATTA	AACGTGCCCC	TTGTTTGTGTT GTTGAGATG 21358
CAGTTTCACT	CTGTCACCCCA	GGCTGAAGTG	CAGTGCAGTG	CAGTGGTGTG ATCTCGGCTC 21418
ACTACAACCT	CCACCTCCCCA	CGTTCAAGCG	ATTCTCATGC	CTCAGCTCC CGAGTAGGTG 21478
GGATTACAGG	CATGACCCAC	TTACACCCGG	CTAATTTTG	TATTTTACTG AGAGCTGGGG 21538
TTTCACCATG	TTGGCCAGGC	TGGTCTCAA	CCCCTAACCT	CAAGTGTATCT GCCTGCCTCA 21598
GCCTCCAAA	CAAACAAACA	ACCCCACAGT	TTAATATGTG	TTACAAACACA CATGCTGCAA 21658
CTTTTATGAG	TATTTTAATG	ATATAGATTA	AAAAGGTG	TTTTTAACTT TTAAATGCTG 21718
GGATTACAGG	CATGAGCCAC	TGTGCCAGGC	CTGAACGTG	TTTTTAAAAA TGTCTGACCA 21778
GCTGTACATA	GTCTCTGCA	GACTGGCCAA	GTCTCAAAGT	GGGAACAGGT GTATTAAGGA 21838
CTATCCTTGT	GTTAAATTTC	CGCAAATGTT	CCTGTGCAAG	AATTCTCTA ACTAGAGTT 21898
TCATTTATTA	TATTTATTTC	AG AT AAT	GCA CCC CGG ACC ATA TTT ATT ATA	21949
		Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile		
40		45		
AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT				21997
Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly				
50	55	60	*	65
GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT				22045
Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile				
70	75			80
TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTCAAT CATGTTAATA TAATCAATAT				22103
Ser Phe Lys				
AATTAGAAAT ATAACATTAT TTCTAAATGTT AATATAAGTA ATGTAATTAG AAAACTCAAA				22163
TATCCTCAGA CCAACCTTT GTCTAGAACAA GAAATAACAA GAAGCAGAGA ACCATTAAG				22223
TGAATACTTA CTAAAAATTAA TCAAAACTCTT TACCTATTGT GATAATGATG GTTTTCTGA				22283
GCCTGTCACA GGGGAAGAGG AGATACAAACA CTGTTTTAT GACCTGCATC TCCTGAACAA				22323
TCAGTCTTA TACAAATAAT AATGTAGAAT ACATATGTGA GTTATACATT TAAGAATAAC				22403
ATGTGACTTT CCAGAATGAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATTT				22463
CTACACCTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTGTG CTGATCCTTA				22523
GCCTAAGTCT TAGACACAAG CTTCAGCTTC CAGTTGATGT ATGTTATTTT TAATGTTAAT				22583
CTAATTGAAAT AAAAGTTATG AGATCAGCTG TAAAAGTAAT GCTATAATTA TCTTCAAGCC				22643
AGGTATAAAG TATTTCTGCC CTCTACTTT TCTCTATTAT TCTCCATTAT TATTCTCTAT				22703
TATTTTCTC TATTCCTCC ATTATGTTA GATAAACCC AATTAACAT ATGCTACAGAC				22763
TGAGCCAGTA AGAGTAGCCA GGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTCC				22823
ATGTGATGAA GACTCTTTT GAGTGGAGAT TTGCAATAA ATATCCGCTT TCATGCCAC				22883
CCAGTCCCCA CTGAAAGACA GTAGGGATAT GACCTTAGTG AAGGTACCAA GGGGCAACTT				22943
GGTAGGGAGA AAAAAGCCAC TCTAAAAT AATCCAAGTA AGAACAGTGC ATATGCAACA				23003
GATACAGCCC CCAGACAAAT CCCTCAGCTA TCTCCCTCCA ACCAGAGTGC CACCCCTCA				23063
GGTGACAATT TGGAGTCCCC ATTCTAGACC TGACAGGCAG CTTAGTTATC AAAATAGCAT				23123
AAGAGGCCCTG GGATGGAAGG GTAGGGTGA AAGGGTTAAG CATGCTGTTA CTGAACAACA				23183
TAATTAGAAG GGAAGGAGAT GGCAAGCTC AAGCTATGTG GGATAGAGGA AAACTCAGCT				23243
GCAGAGGCAATTGATTAGAAC TGGGATAAGT CCGAACCTAC AGGTGGATTC TTGTTGAGGG				23303
AGACTGGTGA AAATGTTAAG AAGATGAAA TAATGTTGG CACTTAGTAG GAACTGGCA				23363
AATCCATATT TGGGGGAGCC TGAAGTTTAT TCAATTTGA TGGCCCTTTT AAATAAAAAG				23423
AATGTGGCTG GCGCTGGTGG CTCACACCTG TAATCCCAGC ACTTTGGGAG GCGGAGGGGG				23483
GCGGATCACC TGAAGTCAGG AGTCAAGAC CAGCCTGAC AACATGGAGA AACCCCATCT				23543
CTACTAAAAA TACAAAATTAA GCTGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG				23603
GGAGGCTGAG GCAGGAGAAT CTTTGAACC CGGGAGGCAG AGGTGCGAT GAGCCTAGAT				23663
CGTGCCTATTG CACTCCAGCC TGGGCAACAA GAGCAAAACT CGGCTCTAAA AAAAAAAA				23723
AAAAAGTGAATTAACAAA GGCATTAGCT TAATAATTAA ATACTGTTTT TAAGTAGGGC				23783
GGGGGGTGGC TGGAAAGAGAT CTGTGAAAT GAGGGAATCT GACATTAAAG CTTCATCAGC				23843
ATCATAGCAA ATCTGCTTCT GGAAGGAAC CAATAAAATAT TAGTGGAGG GGGGGAGAGA				23903
GTGAGGGGTG GACTAGGACC AGTTTAGCC CTTGTCTTTA ATCCCTTTTC CTGCCACTAA				23963



GTAGTGACCC	ATGCCCTCAA	TCCCAGCTAC	TCAGAGGCT	GAGGCAGGAG	AATCACTTGC	27267
ACTCCGGAGG	TGGAGGTTGT	GGTAGGCCGA	GATTGCACCA	TTGCGCTCTA	GCCTGGGCAA	27327
CAACAGCAAA	ACTCCATCTC	AAAAAATAAA	ATAAATAAAT	AAACAAATAA	AAAATTCTATA	27387
ATGTGAACTG	TCTGAATTTC	TATGTTAGA	AAGATTATGA	GATTATTAGT	CTATAATTGT	27447
AATGGTGAAA	TAACATAAAT	ACCAGTCTG	AAAAACATCA	TTAAGAAATG	AATGAACCTT	27507
CACAAAAGCA	AAACAAACAGA	CTTATCCCTA	TTAAGTGAA	AAAATAAAA	AAAATAAAA	27567
TAATGTTAA	AAAATTCTATA	GTGGAAAAC	ATTCTACATT	GTTAATTGGC	ATATTAAATTA	27627
TACTTAATAT	AATTATTTT	AAATCTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG	27687
ATATTTGAAAC	TTTCTAATTTC	TTAAGAATAT	CGTTAACCA	TCAATATTTC	TATAAGGAGG	27747
CCACTTCACT	TGACAAATTTC	CTGAATTCC	TCCAAAGTCA	GTATATTTC	AAAATTCTAGT	27807
TTGATCCTGA	ATCCAGCAAT	ATATAAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCAT	27867
CCTAGGAATG	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AAACATAACAT	ACTATATTAA	27927
TAAAGTATCA	AAACAGTATT	CTCATCTTT	TTTCTTTTT	CACAATTCC	TGGTTACACT	27987
ATCATCTCAA	TAGATGCAGA	AAAAGCATTT	GACAAAATCC	AATTCTATAAT	AAAATTCTC	28047
AAACCTGAAA	GAGAACATCA	TAAAGGCATC	TATGAAAAC	CTACAGCTAA	TATCATACTT	28107
AACGATGAAA	AACTGAATT	TTTACCCCTA	AGATCAAGAA	TAATGCAAGC	ATGTCAAGCTC	28167
TTGCAACTC	TATTCAACAT	TGTACTGGAG	GTCTCTAGCCA	GAGCAACCAT	ACAATAAAATA	28227
AAAATAAAAG	GCACCCAGAT	TAGAAAGGAA	GTCTTTATTTC	GCAGACAAACA	TGGTTCTTA	28287
TGCAGAAAAC	CGTCAGGAAT	ACACACACAT	GTAGAAGTCA	ATAAGGTCAG	CAAGGTTGCA	28347
GGTTGCAATA	TCAATATGCA	AAAATACATT	GAAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
CCTTCGTCC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTG	28467
AGGCTATAGT	GCAATGTGAT	CTTGCCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA	28527
AAGTGAGACC	CCGTCTCCAA	AAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT	28587
GAATGATCTG	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	AAAAAAATA	AAATACAAAT	28647
AAATTTAGCA	AAATAATTAT	AAAACTTGTA	CATCGAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
ATTAAAGATG	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATTGG	AAAATTCTATT	28767
CAATATTGTT	AAGATAACAA	TTGTCCCCAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT	28827
CAAATTCCA	GCAGGGTTT	TGCAGAAATT	GACAAGCTGT	ACCCAAAATG	TATATGGAAA	28887
TGAAAAGACC	CAGAAGAGCA	AATAATTTC	AAAAACAAA	GTTGGAAAAC	TTTACTTCC	28947
TAATTTAAA	ACTTACTATA	AACCTAAAGT	TATCAAGAGCC	ATTTAGT		28994

(15) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (v) FRAGMENT TYPE: N-terminal fragment
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser  
 1                   5                   10

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 27 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGGT

18

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 28 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC

30

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC

30

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTCAT TGCCACAAAG TTG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCAAGT AGAAGACAAT TGC

33

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTGGTC

27

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAATCA GATTACTTG GCAAGCTTGA ATC

33

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGGT TG

32

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAACTATCT C

31

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG

33

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTCAG AGAAGTGTCC

30

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTGAA TCTTCATCAT ACCAAGGATA C

31

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCAGT A

31

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTAATA GCTAGTC

37

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAATCA GATTACCTTG GCAAGCTTGA ATC

33